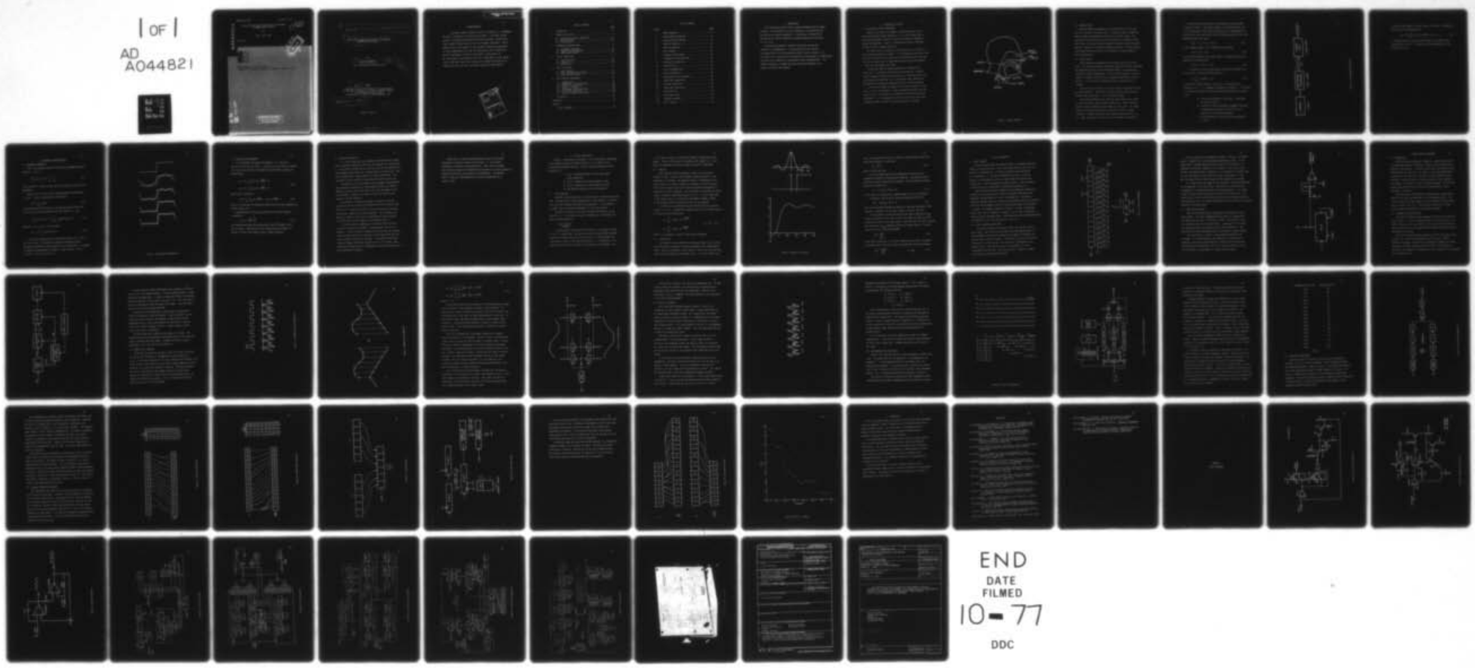
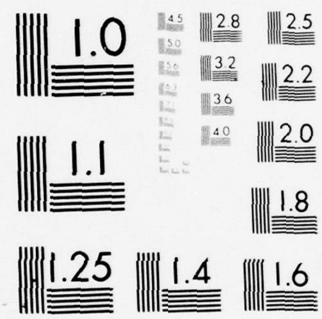
AD-A044 821 ILLINOIS UNIV AT URBANA-CHAMPAIGN DEPT OF COMPUTER SCIENCE F/G 17/2
APPLICATION OF BURST PROCESSING TO THE SPECTRAL DECOMPOSITION O--ETC(U)
JUL 77 C J XYDES N00014-75-C-0982
UNCLASSIFIED UIUCDCS-R-77-870 NL

| OF |
AD
A044821



END
DATE
FILMED
10-77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UIUCDCS-R-77-870 UILU-ENG 77 1745

ADA044821

(12)

APPLICATION OF BURST PROCESSING TO THE SPECTRAL DECOMPOSITION OF SPEECH

by

CHRIST JOHN XYDES

uly 1977

DDC
OCT 4 1977
C

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN · URBANA, ILLINOIS

AD No.
DDC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

(14) UIUCDCS-R-77-870, UILU-ENG-77-1745

(6) APPLICATION OF BURST PROCESSING TO THE SPECTRAL DECOMPOSITION OF SPEECH

BY

(10) CHRIST JOHN XYDES

B.S., University of Illinois 1975

(11) Jul 77

(12) 62p.

(9) Masters THESIS,

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1977

(15) N00014-75-C-0982

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Urbana, Illinois

176 011

mt

PRECEDING PAGE NOT FILMED
BLANK

ACKNOWLEDGEMENT

The author wishes to thank his advisor, Professor W. J. Poppelbaum, for year of friendship and professional guidance. It has been a great pleasure to have been associated with him and his group. He would also like to thank Professor Jane Liu for her friendship and advice. Special thanks are due to the members of the Information Engineering Laboratory for their companionship over the last three years; to Mr. Frank Serio and Mr. Sam McDowell for help in the project's construction; to Mr. Stan Zundo for the drafting; and to Ms. Cinda Robbins for the typing. Finally, the author would like to express his deepest gratitude to his parents and sister for their continuous support and understanding.

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
A

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

This investigation deals with the spectral decomposition of speech waveforms. The motivation for such an operation is the applicability to areas such as speech compression. A large body of references on applications of various transforms to speech processing can be found. [9, 10, 11, 15]

The major shortcoming of transform processing has been the complexity of implementation. A unique solution to the problem is proposed which utilizes advantages present in Burst Processing. [3] The feasibility of using such an unconventional representation is demonstrated and shown to be preferable to conventional binary implementations. The inherent properties of speech have been exploited throughout in an attempt to minimize the hardware.

2. PROPERTIES OF SPEECH

2.1 PHYSIOLOGY OF SPEECH PRODUCTION

Speech is the result of voluntary, formalized motions of the respiratory and masticatory apparatus. It is a skill which must be learned and developed. Control is aided by the acoustic feedback of the hearing mechanism. Figure 1 illustrates the parts of the human anatomy relevant to speech production.

The vocal tract is an acoustical tube which acts as a filter on the excitation functions of speech. It is terminated by the lips on one end and by the vocal cords at the top of the trachea on the other end. The cross sectional area is nonuniform and may be varied by movement of the lips, jaw, tongue, and velum.

An ancillary path for speech production is provided by the nasal tract. It extends from the velum to the nostrils. Acoustic coupling between the nasal and vocal tracts is controlled by the size of the opening at the velum. As is well known, nasal coupling can substantially influence the characteristics of the sound produced.

The source of energy for speech lies in the air flow out of the lungs. As air is forced out, it passes through the trachea into the throat cavity. At the top of the trachea one finds the vocal cords and glottis. It is the degree of activity of the vocal cords which determines whether "voiced" or "unvoiced" speech is produced.

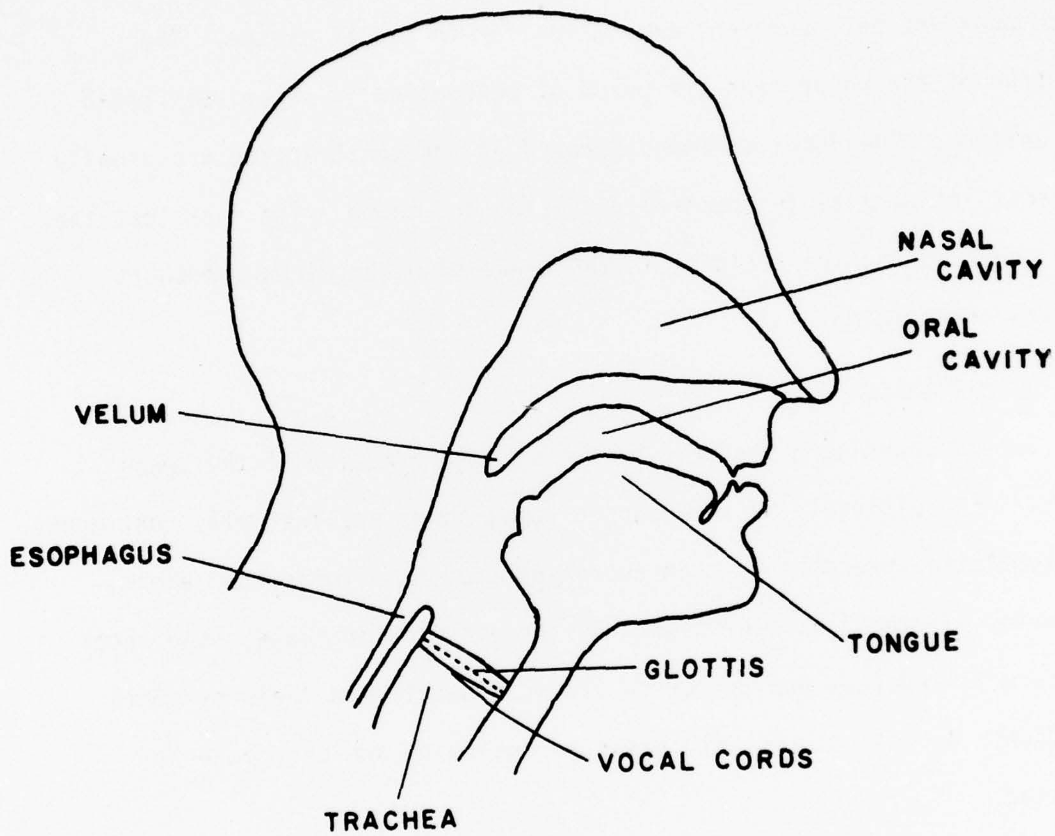


Figure 1. Speech Apparatus

2.2 UNVOICED SPEECH

Unvoiced sounds are produced by a turbulent flow of air at some point of stricture in the vocal tract. An acoustic noise is generated which provides an incoherent excitation for the vocal system. The spectrum of the noise near its point of generation is relatively broad and uniform. The vocal cavities forward of the construction are usually the most influential in spectrally shaping the sound. The fact that the vocal cords do not participate in the creation of unvoiced speech is the key observation.

2.3 VOICED SPEECH

Voiced sounds are produced by the vibratory action of the vocal cords. The relatively massive tensed vocal cords are initially contiguous. The subglottal pressure is then increased enough to force them apart, producing a lateral acceleration. As the air flow increases, the local pressure is reduced, and the cords are returned toward their original position. As this occurs, the pressure builds up and the cycle is repeated.

The period of oscillation of the vocal cords is determined by their mass and compliance. This period is usually shorter than the natural period of the cords; thus, it is a forced oscillation.

The orifice produced by the vibration cords breaks up the steady air flow into short, quasi-periodic pulses of air. These pulses are used to excite the acoustic system above the vocal cords. The volume flow of air through the glottis as a function of time is roughly triangular in shape and exhibits duty factors on the order of 0.3 to 0.7. Thus, the glottal air flow is rich in harmonics and overtones.

A simplified block diagram for the production of voiced sounds is shown in Figure 2. The output signals $S_v(t)$ appearing at the lips is the convolution of the excitation function e(t), corresponding to the air flow at the vocal cords, with the impulse response of the filter representing the vocal tract.

$$S_v(t) = \int_{-\infty}^{t} e(t)\ v\ (t-k)\ dk \tag{2.1}$$

In the frequency domain, this corresponds to the product

$$S_v(f) = E(f) \cdot V(f) \tag{2.2}$$

The amplitude spectrum of the speech signal is obtained by taking the magnitudes of the functions.

$$|S_v(f)| = |E(f)| \cdot |V(f)| \tag{2.3}$$

This process may also be considered from a Fourier decomposition point of view. Writing the source signal as

$$C_v = \sum_{h=1}^{H} A_h \cos(hFt + \theta_h) \tag{2.4}$$

we consider H audible harmonics, each with its own amplitude A_h frequency hF (F = 1/T - fundamental frequency), and phase θ_h. Information is transmitted through the following modulation processes of the vocal tract:

1) Starting and stopping of the source - represented by the function s(t).
2) Variation of the instantaneous fundamental frequency represented by replacing Ft with $F\int_0^t i\ (t)\ dt$, where i(t) is the inflection factor.
3) Filtering effects of the vocal tract represented by v(t).

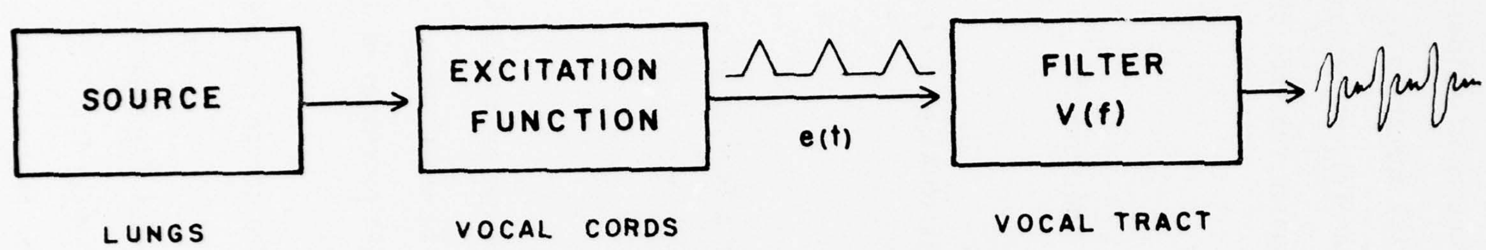


Figure 2. Voiced Speech Production

In normal voiced speech, all three factors are present simultaneously, giving a wave form represented by

$$S_v(t) = s(t) \sum_{h=1}^{H} v(t)\, A_h \cos(hF\int_0^t i(t)\, dt + \Theta_h) \qquad (2.5)$$

As stated previously, the amplitude spectrum of the speech signal, represented by $|S_v(f)|$, is obtained by taking the magnitude of the transform of $S_v(t)$.

3. ORTHOGONAL REPRESENTATIONS

3.1 ORTHOGONAL EXPANSIONS

A set p of arbitrary functions is said to be orthogonal over the interval $t_1 < t < t_2$ if

$$\int_{t_1}^{t_2} P_i(t)\, P_j(t)\, dt = \begin{matrix} c & i = j \\ 0 & i \neq j \end{matrix} \tag{3.1}$$

If the constant c is equal to one, the set of functions is said to be orthonormal.

Suppose $S_v(t)$ is a real valued function defined on the interval (t_1, t_2). It can be represented by the expansion

$$S_v(t) = \sum_{h=0}^{\infty} a_h P_h(t) \tag{3.2}$$

To evaluate the k^{th} coefficient a_k, one multiplies both sides of Eq. (3.2) by $P_k(t)$ and then integrates over the interval (t_1 , t_2).

$$\int_{t_1}^{t_2} S_v(t)\, P_k(t)\, dt = \int_{t_1}^{t_2} \sum_{h=0}^{\infty} a_h P_h(t)\, P_k(t)\, dt \tag{3.3}$$

Applying Eq. (3.1) to Eq. (3.3) we obtain

$$a_h = 1/c \int_{t_1}^{t_2} S_v(t)\, P_k(t)\, dt \tag{3.4}$$

$S_v(t)$ may be approximated by limiting the series in Eq. (3.2) to the first H terms. The amount of distortion introduced by this approximation depends on the characteristics of the function. An example of converging approximations is shown in Figure 3 with P = [sin x, (sin 3x)/3, (sin 5x)/5, (sin 7x)7].

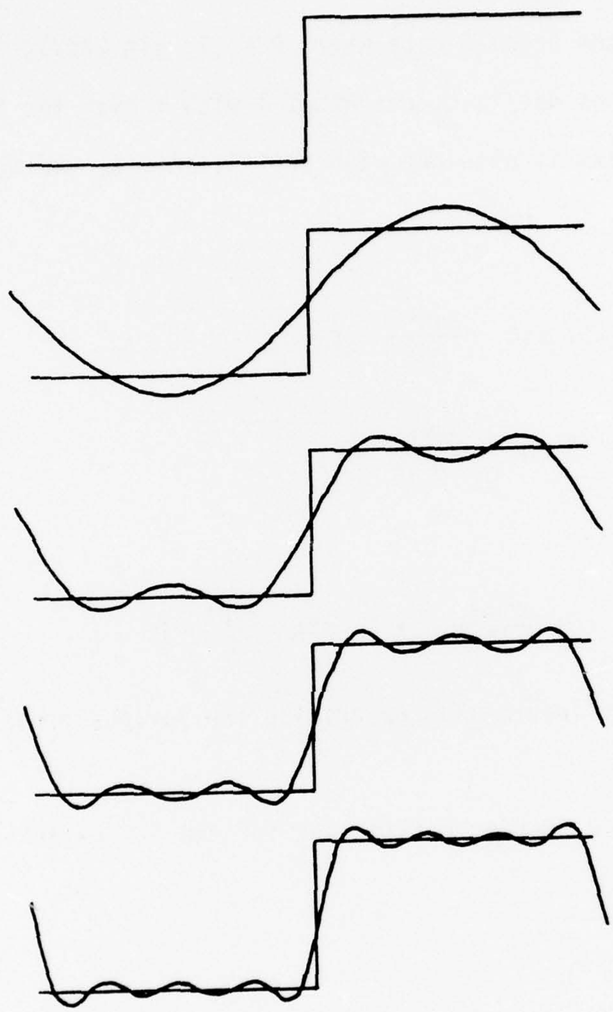


Figure 3. Converging Approximations

3.2 FOURIER SERIES EXPANSION

If one considers the special case where P = [1, sin 2Δt/T, cos 2Δt/T, sin 4Δt/T, cos 4Δt/T,... cos hΔt/T 1 with t over the interval [0,T], the Fourier Series is obtained with coefficients a_h and b_h as defined below.

$$a_h = 2/T \int_0^t S_v(t) \sin \frac{2\pi ht}{T} dt$$

$$b_h = 2/T \int_0^t S_v(t) \cos \frac{2\pi ht}{T} dt \tag{3.5}$$

where $S_v(t)$ is defined as

$$S_v(t) = \sum_{h=1}^{\infty} [b_h \cos \frac{2\pi ht}{T} + a_h \sin \frac{2\pi ht}{T}] \tag{3.6}$$

There is a large body of information describing the various properties of Fourier Series. [14]

The magnitude of the Fourier coefficient for the h^{th} harmonic is defined as

$$|S_v(h)| = \sqrt{a_h^2 + b_h^2} \tag{3.7}$$

This quantity describes the contribution of the h^{th} harmonic to the overall signal. Application of such information can be found in a variety of fields, most notably that of signal processing.

3.3 COMPRESSION PROPERTY

It is a well known fact that orthogonal transformation of signals offer a potential reduction in the bit rate necessary for transmission. [1] This ability follows directly from the fact that the magnitudes of the orthogonal coefficients are a strong function of their order. Most of the information is concentrated in the lower coefficients. Thus the number of bits required to represent each coefficient can vary.

If compression is to be achieved using orthogonal transformations when compared to standard pulse code modulation (PCM), the average number of bits per coefficient must be less than the number of bits per PCM sample. However, the signal to noise ratio must be kept constant to allow comparison.

The necessary relationships between the number of bits per coefficient and the variance of the coefficient have previously been derived. [1,2] Once the variances of the coefficients are measured, the required number of bits can be computed. This was done using samples of speech for the first 16 coefficients of three different transformations--Fourier, Hadamard, and Karhunen-Loeve. [1]

The results may be summarized by listing the transformations in order of decreasing performance: Karhunen-Loeve, Fourier, and Hadamard. To equal the SNR of standard 56 Kbit/sec PCM, the Karhunen-Loeve transform required 42.5 Kbit/sec, Fourier required 46 Kbit/sec, and the Hadamard required 48.5 Kbit/sec. The maximum difference of 6 Kbit/sec between transforms represents only a 14% increase. When considering the complexity of implementation, one might consider such a small degradation acceptable.

Comparisons of signal-to-quantizing-noise ratios for the three transforms at various bit rates are also known. [1] The relative performance is the same as observed earlier. Thus, one can conclude that bit-rate savings can be achieved at the expense of increased processing complexity required by the orthogonal transformations. An alternate result is that such transformations will allow increased SNR for a fixed bit rate.

4. DIGITAL COMPUTATION

Often it is desirable to evaluate Eq. (3.7) using digital technology. If one chooses to enter the digital domain, Eq. (3.5) and Eq. (3.7) can never be evaluated precisely. The main factors preventing infinite precision are

1) $S_v(t)$ is observed through a finite time window (time truncation)
2) $S_v(t)$ is sampled at discrete instants in time
3) $S_v(t)$ is quantized to a fixed number of levels
4) Any machine possesses only finite precision

4.1 TIME TRUNCATION

A machine can only deal with a finite portion of a signal at any given time. Considering this window through which it sees the world, its effect is to limit the frequency resolution of the analysis. If the window is T seconds long, only spectral components 1/T Hz can be resolved.

The Fourier transform of the unit amplitude data window is of the form sin x/ x. If a sinusoidal input of frequency f_0 is considered, the spectrum obtained would be

$$\frac{\text{Sin}\ (\pi(f-f_0)T)}{\pi(f-f_0)T}$$

In the processor to be described in section 6, a rectangular window of size T is always used. One can use this fact to analyze the amplitude distortion introduced on the resultant spectrum. If one considers the input speech to have a flat spectrum starting at F, a convolution of this spectrum with the sin x/x response of the window can be performed. The

result, shown in Figure 4, indicates the amount of distortion one may expect. Since this distortion is stationary with respect to F, it can easily be compensated for before further processing is undertaken.

4.2 SAMPLING

Ideal sampling involves observing a signal only at discrete instants in time. Usually, these samples are equally spaced in time - separated by Δt seconds. The sampling function is represented in the Fourier domain as a train of impulses of strength Δt, each $1/\Delta t$ apart. Multiplying the input signal by the sampling function corresponds to a convolution of their transforms. This will amount to repeating the input signal's transform around each multiple of $1/\Delta t$. It is well known that if the sampling rate is at least twice the highest frequency present in the original signal, so-called aliasing will be prevented.

In section 3.2 we defined the Fourier coefficients for continuous signals - Eq. (3.5). Considering a sampled signal, one can define the pair of Discrete Fourier Coefficients

$$a_h = \sum_{k=0}^{d-1} S_v(k\Delta t) \sin \frac{2\pi hk\Delta t}{T}$$

$$\Delta t = T/d \qquad (4.1)$$

$$b_h = \sum_{k=0}^{d-1} S_v(k\Delta t) \cos \frac{2\pi hk\Delta t}{T}$$

where d is the number of points in the discrete transform.

4.3 QUANTIZATION

To represent an input sample with continuous values using a finite precision machine, the sample value must be mapped into a representable value. The noise introduced in this process is due to the fact that many input values are mapped into one output value. It is a well known fact

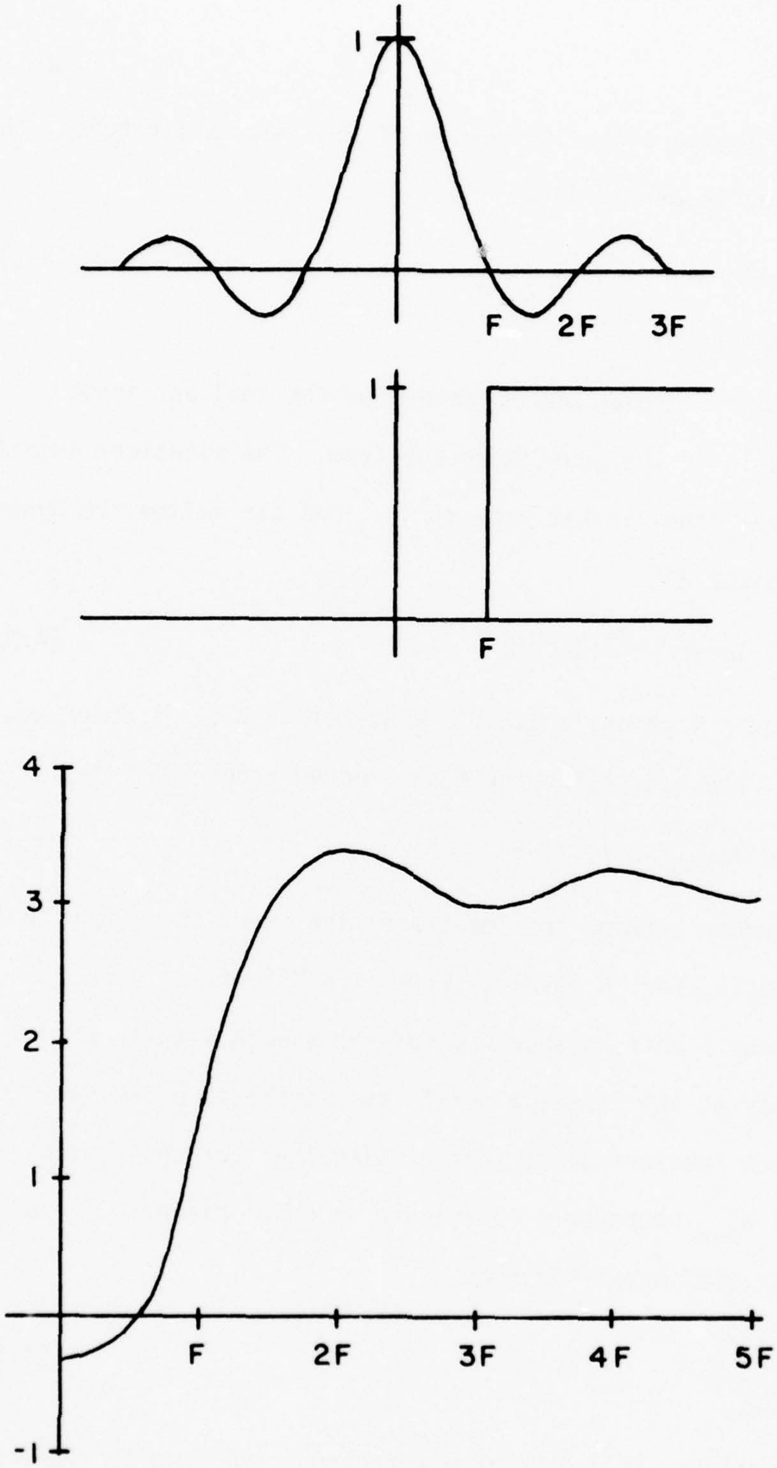


Figure 4. Spectral Distortion

that if the quantization error is treated as zero mean white noise, the noise power produced is of the form

$$N_q = q^2/12 \tag{4.2}$$

where q is the step size.

This noise power enables one to determine the smallest input component discernable in the quantizers spectrum. The quantizer output must have a spectral density that exceeds Nq. One can define the Dynamic Range of the quantizer as

$$DR = -10 \log_{10} [q^2/12] \text{ db} \tag{4.3}$$

For 16 levels of quantization, a Dynamic Range of 34.9 db is observed.

Alternately, one can define the mean squared error (MSE) as

$$MSE = 10 \log_{10} [q^2/12] \text{ db} \tag{4.4}$$

This is the difference between the spectra of the input and output of the quantizer. Obviously, for 16 levels, there is a MSE of -34.9 db.

If one assumes a uniform spectrum for the speech signals with H_{max} harmonic components at the input, a worst case signal to noise ratio can be derived for each Fourier coefficient. Using the fact that q=1/b, consider the case when all H_{max} components contribute an equal amount. The signal to noise ratio per coefficient becomes

$$SNR = \frac{12b^2}{H_{max}^2} \tag{4.5}$$

Under these assumptions, if H Fourier coefficients are used to represent the speech signal, the total signal to quantization noise ratio becomes

$$SNR_T = \frac{12Hb^2}{H_{max}^2} \qquad H \leq H_{max} \tag{4.6}$$

5. BURST PROCESSING

5.1 BURST CONCEPTS

It has previously been the accepted practice to represent quantized signals as binary data words. Such a PCM scheme requires log b bits for be levels of quantization. In 1974, an alternative was proposed by W. J. Poppelbaum. [3] Instead of representing b levels in a binary fashion, it was proposed to utilize a unary scheme and represent the b levels with b equally weighted bits (Burst digits). Such a reduction in precision may be counteracted by appropriate averaging. [16]

During the past two years, members of the Information Engineering Laboratory of the University of Illinois have been investigating the properties and applicability of such a representation. Designed as a compromise between stochastic processing and weighted binary, Burst exhibits simplicity and acceptable accuracy for applications where time averaging is allowed. The hardware complexity of Burst is an order of magnitude greater than that of stochastics. Howver, it is an order of magnitude less than that of weighted binary. Applicable areas include AM demodulation, FM demodulation, and video transmission. [4,5,6,7,8]

5.2 BURST ENCODING AND DECODING

The digital encoding of an analog signal into the Burst domain is quite simple. Many variations of encoders have been demonstrated. [8] The fundamental building block common to all schemes is the Block Sum Register (BSR), shown in Figure 5. Consisting of a b-bit shift register connected to b current sources, this particular implementation uses negative logic. Each current source is activated by a 0 in the corresponding bit position. The total current is summed on a common bus producing a quantized-analog output.

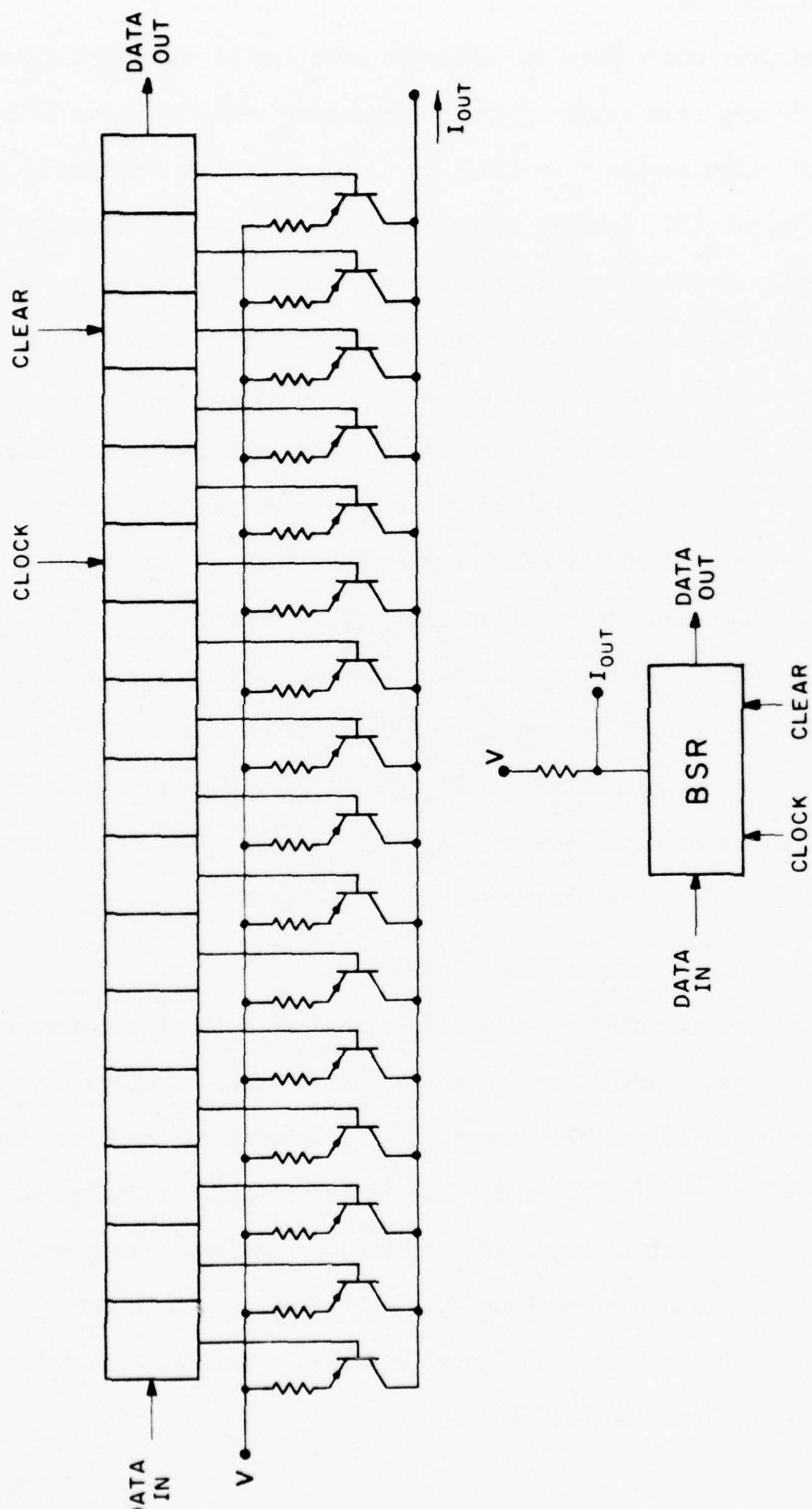


Figure 5. Block Sum Register

A Burst encoder may be implemented as shown in Figure 6. The analog signal is compared to a staircase waveform generated by a BSR. If the analog input is greater than the present value of the staircase, a 1 is produced at the output; otherwise a 0 is produced. Thus, after b clock periods, a new Burst sample is produced. It is compacted in the sense that all ones are adjacent to each other at one end of the sample. If the BSR uses negative logic, the two inputs of the comparator are switched.

It is obvious that the number of ones produced is directly proportional to the magnitude of the input signal. The step size q of the staircase is dependent on the maximum amplitude of the analog signal. It is chosen so that the peak-to-peak variation of the input rarely exceeds bq. The effects of not using a sample-and-hold at the signal input have previously been discussed. [5,8] For improved performance, one may elect to use a sample-and-hold at the analog input.

5.3 BURST MULTIPLICATION

Burst multiplication may be implemented in the digital or quasi-analog domain. The latter implementation was chosen for reasons which will become obvious later. Referring to Figure 5, the voltage V serves as a weighting factor for the stored Burst. Increasing V will increase the quantized analog value present on the current summing bus. Thus, multiplication can be performed without any increase in digital hardware.

This key result is critical to the hardware realization to be presented. It is well known that the complexity of conventional FFT processors using binary representation is largely due to the required multiplications and additions. [12] It will be shown that Burst allows such operations to be performed in a highly parallel manner.

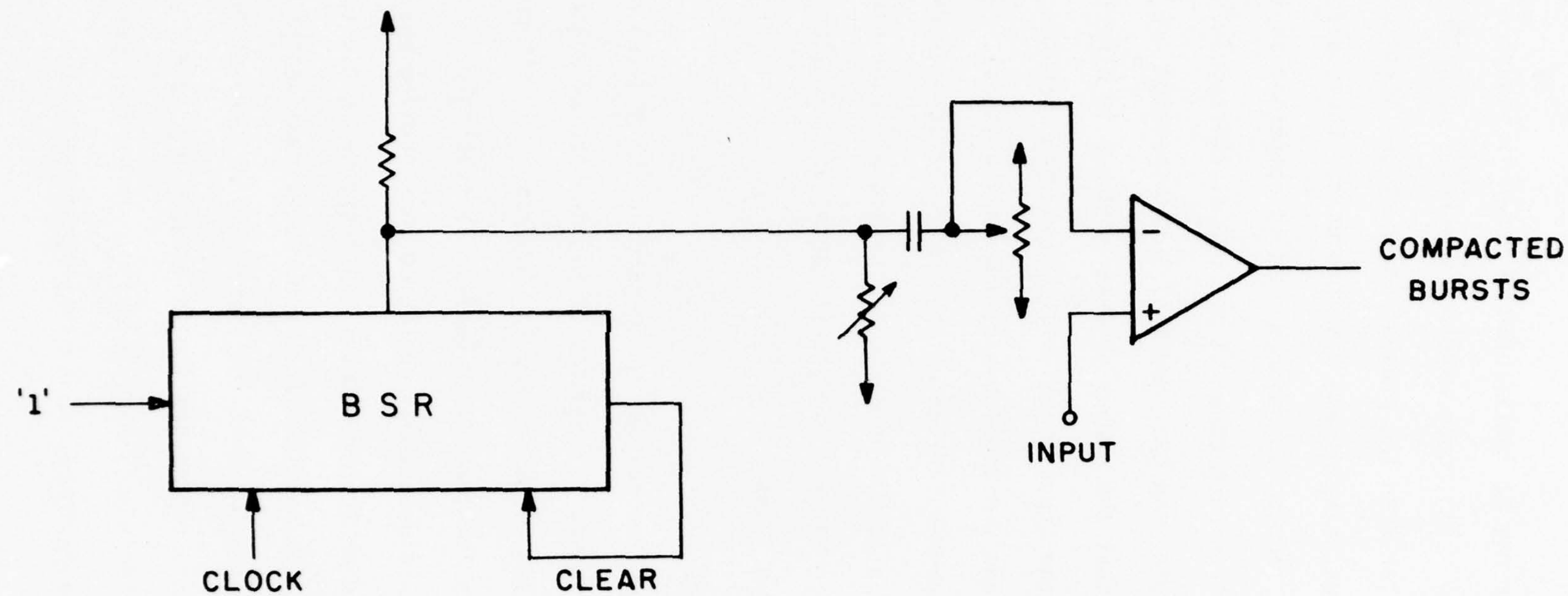


Figure 6. Burst Encoder

6. BURST FOURIER TRANSFORMER

6.1 INTRODUCTION

Given the previous background information, a detailed description of the prototype machine is possible. Figure 7 shows a general block diagram of the processor. The speech signal enters an analog front end which performs two functions. The signal is initially passed through an amplifier with a gain of 2.5 to obtain a signal capable of being processed. Since the signal is locally accessible, it was decided to use automatic gain control instead of adaptive encoding. The amplified signal enters an AGC circuit and also the pitch detection circuit described in section 6.2.

The Asynchronous Pulse Multiplier (APM) generates the appropriate sampling clock given the beginning of each fundamental period. This clock is used to drive the transform unit which performs the multiplications indicated in Eq. (4.1). The resulting coefficients are then used to compute the magnitude of the spectral component.

6.2 FUNDAMENTAL PERIOD DETECTION

The problem of detecting the fundamental pitch period of speech is highly complex. In fact, a complete solution is yet to be found. The main difficulty is that voice pitch is not a clearly defined attribute. Precisely what epochs of the speech waveform should be chosen for period measurement is not clear.

Most pitch extraction methods attempt to identify the epoch of each glottal puff. Describing the periodicity of the signal, inverse filtering techniques, or measuring the fundamental component are common approaches. The most promising of these is the so-called cepstrum technique. [10] However, the complexity of such an approach is overwhelming for many applications.

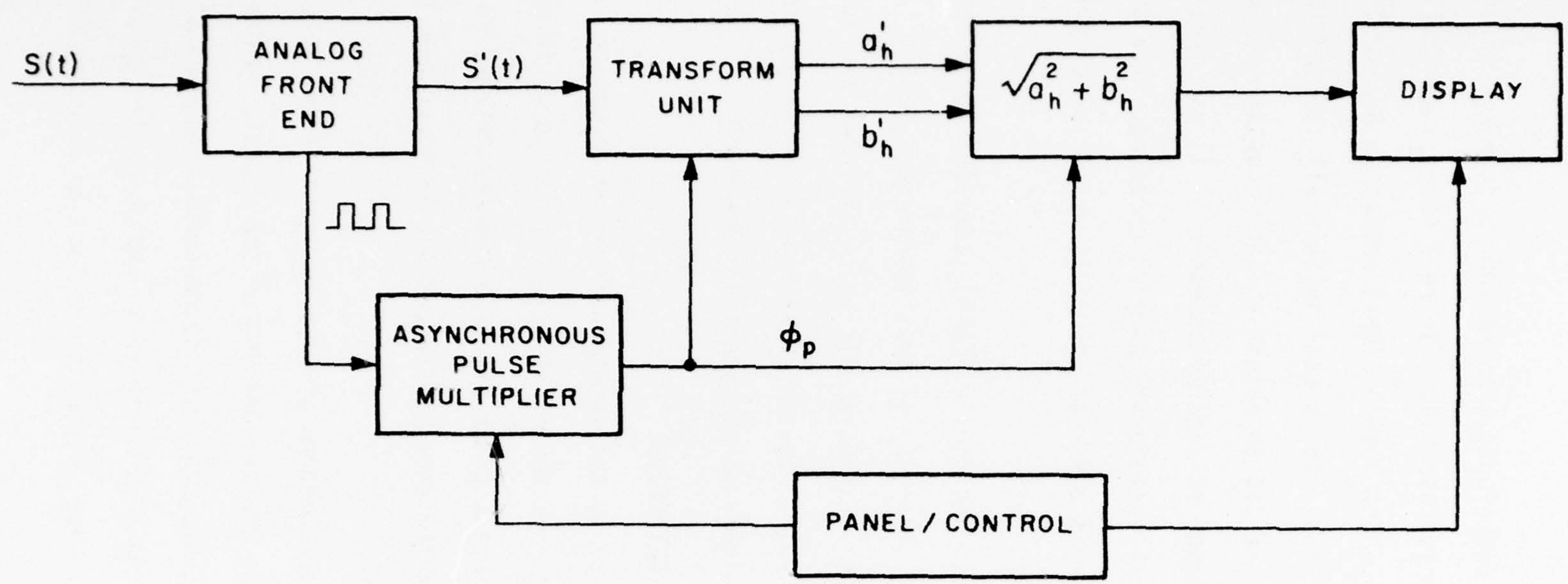


Figure 7. Processor Block Diagram

The pitch detection method implemented takes advantage of the rapid initial rise of the speech waveform. A search waveform performs a linear search for the speech peak. In order to remove the effects of transients, the search waveform is reset after a delay of 0.7 ms. The reset level is set to a fixed quantity above the peak of the speech. Thus, the search waveform will track the speech amplitude.

At the point of each detection, an output pulse is generated to signify the beginning of the pitch period. A pulse duration of 2.4 ms is used to mask off possible retriggering by peaks within the same fundamental period. Such a mask performs the function of a low pass filter with a cutoff frequency of 417 Hz. Figure 8 shows an actual trace of the circuit in operation.

Personal observations indicate a high degree of tracking. The problems which may be introduced by the changing phase of the signal may cause a frequency modulation effect. Although this may be slightly bothersome, it will not prevent intelligibility.

6.3 HARMONIC SELF SAMPLING

The problem of convolving the speech signal $S_v(t)$ with sin hwt and cos hwt for the h^{th} harmonic is fundamental to the calculation of $|S_v(t)|$. An alternate approach is the idea of Harmonic Self Sampling. [17] If one divides a given period of $S_v(t)$ into h+1 equal segments, only one period of a sine and cosine waveform is required. One merely performs h+1 partial convolutions of each segment with the sine and cosine. Summing over these partial convolutions and scaling appropriately, one obtains the coefficients a_h and b_h. This is illustrated in Figure 9. Using this idea, Eq. (4.1) now becomes

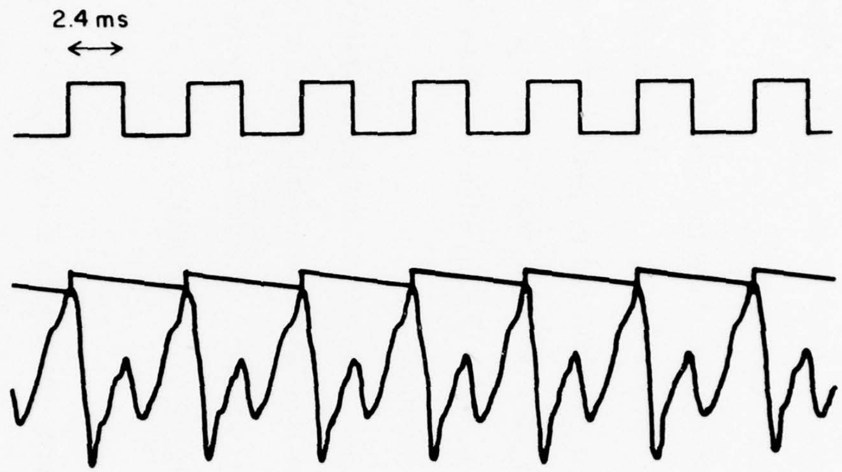


Figure 8. Fundamental Period Detection

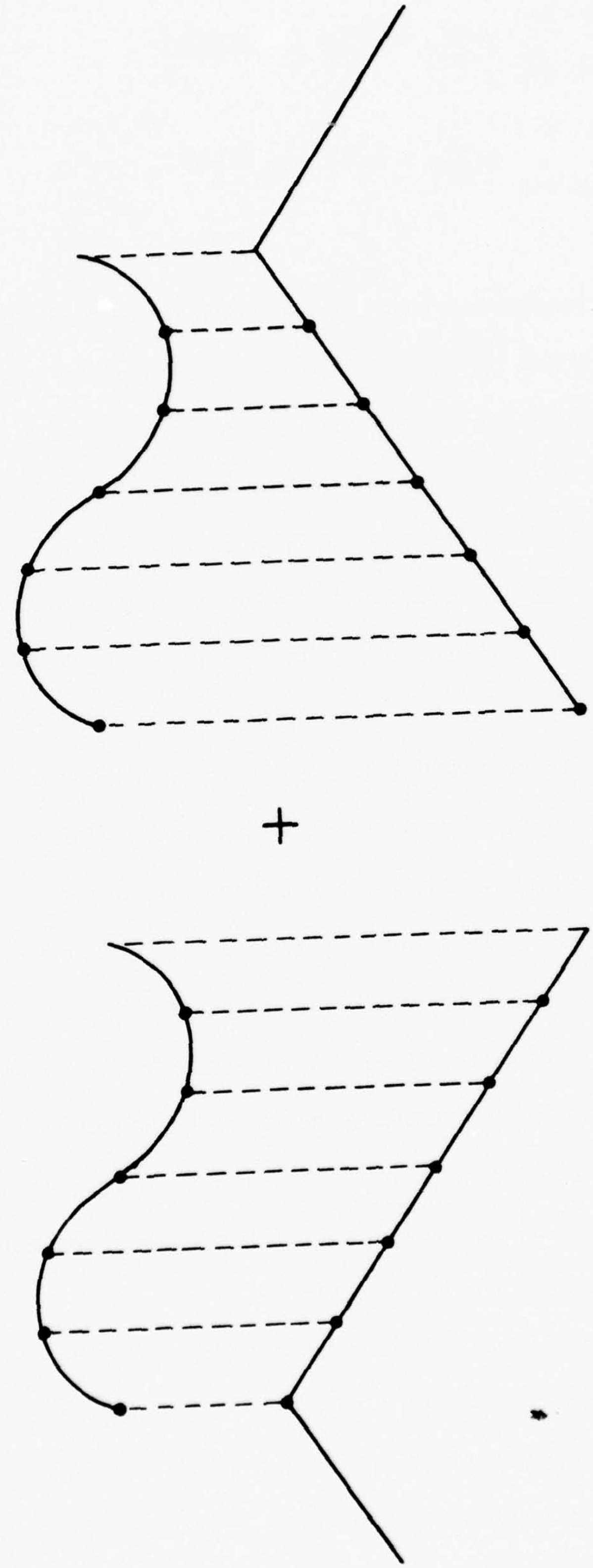


Figure 9. Harmonic Self Sampling

$$a_h = \frac{1}{h+1} \sum_{n=0}^{h} \sum_{k=0}^{d-1} S\left(\frac{nT}{h+1} + \frac{k\Delta t}{h}\right) \sin \frac{2\pi k\Delta t}{T}$$

$$b_h = \frac{1}{h+1} \sum_{n=0}^{h} \sum_{k=0}^{d-1} S\left(\frac{nT}{h+1} + \frac{k\Delta t}{h}\right) \cos \frac{2\pi k\Delta t}{T} \quad (6.1)$$

where $\Delta t = T/d$.

The motivation behind such an approach is that the weighting voltages on a row of BSR's can be adjusted to simulate a given waveform. The transform unit, shown in Figure 10, consists of two rows of 32 BSR's, one row weighted with a sine wave, the other weighted with a cosine wave. By adjusting the input sampling rate appropriately, these voltages remain stationary. Each Burst encoded subsection of speech is passed through these two rows. After the complete subsection is present, the current output is observed.

Using this implementation, the hardware complexity of standard Fourier transformers is circumvented. Two rows of BSR's with appropriate voltage sources replace the required complex arithmetic units. Storage elements are required independent of the type of processing techniques implemented. Using weighted binary, each of the registers requires log b bits. However, additional storage is required for the complex constants involved. The indexing and control hardware required for the complex arithmetic unit must also be considered. [12] In comparison, the increase in hardware needed to perform the required convolutions in the Burst implementation is almost negligible.

Due to these parallel multiplications and additions, the number of computations is also reduced to a minimum. With regard to Eq. (6.1), the inner summation is performed in one step. Thus, there are order of h computations for the h harmonic and total of (H+1)(H+2)/2 computations for a complete spectrum of H+1 harmonic lines.

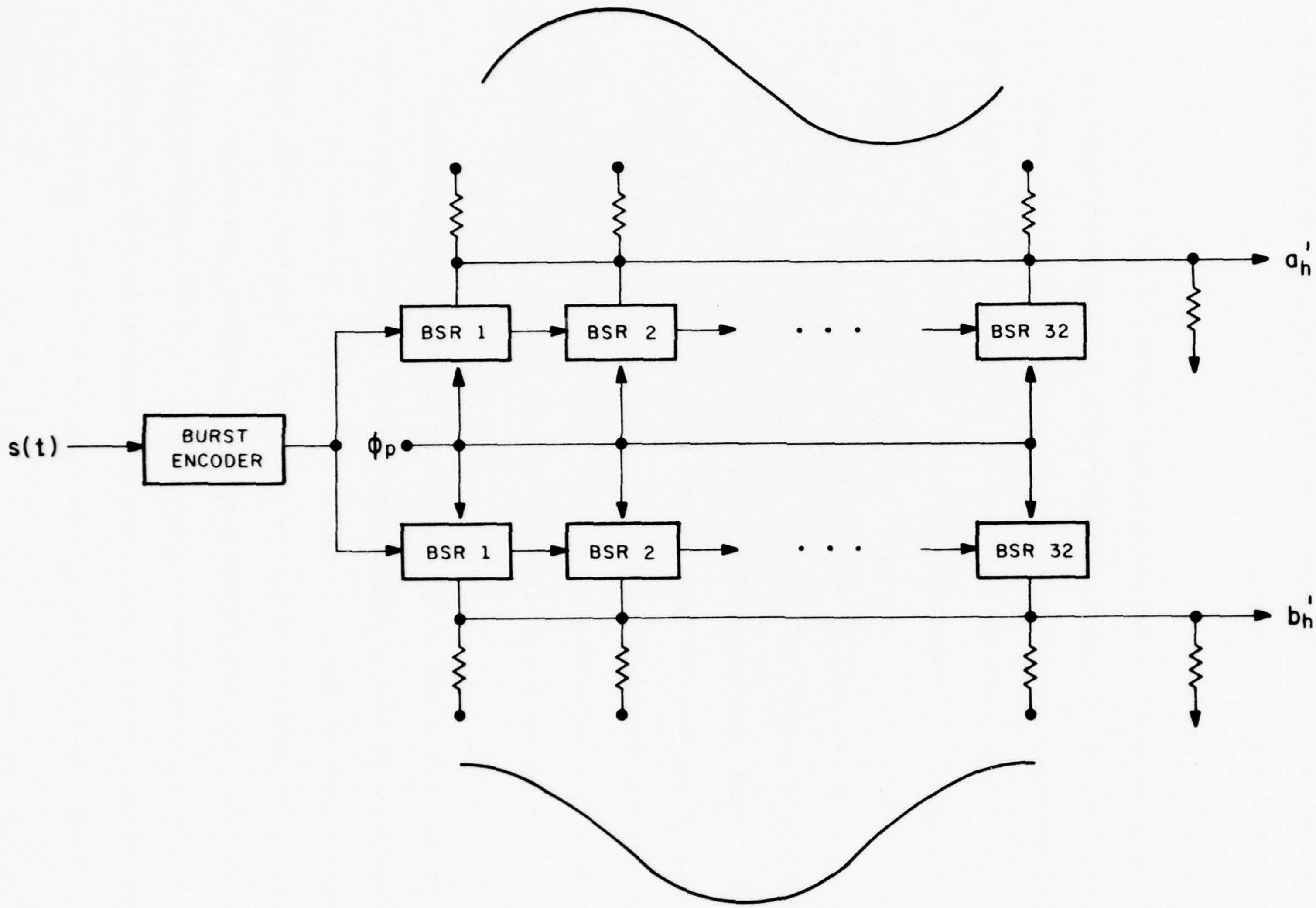


Figure 10. Transform Unit

The time delay involved in the calculation approaches zero. As data serially enters the processor, the required partial convolutions are performed on-line and the results are accumulated. After the final convolution, $|S_v(h)|$ is computed. The time required for this computation is the total delay encountered.

6.4 SERIAL VS. PARALLEL

Due to the highly redundant nature of speech, if one is only interested in a small number of coefficients, a single coefficient may be computed each fundamental period. For H coefficients, this would require H periods, as shown in Figure 11. Assuming the use of a d-point transform unit, with each point consisting of b bit Bursts, we obtain the following results. For a given harmonic h (h=0 to 7), the fundamental period T is divided into bd(h+1) samples. Thus, the input sampling rate is (bd(h+1))/T samples per second.

The output, consisting of a number of spectral lines (8 in this implementation), is pitch synchronous. Using a range of 50 Hz to 250 Hz for the fundamental period, one obtains a rate of 6.25 spectra per second to 31.25 spectra per second. This corresponds to 800 to 4000 Burst digits per second, or an equivalent 200 to 1000 binary digits per second.

If one rejects the serial approach, a parallel analysis may be implemented. The idea of partial convolutions can still be used at the expense of added hardware. If one is interested in the first H+1 harmonics, H+1 data streams must be maintained in parallel. This implies H+1 transform units, coefficient computation hardware, and APM's.

A more subtle approach is also possible. Fix the input sampling rate at (bd(h+1))/T. Thus, a_H and b_H are obtained directly from the sampled input stream. To obtain the a_h's and b_h's for h=0 to H-1, one may

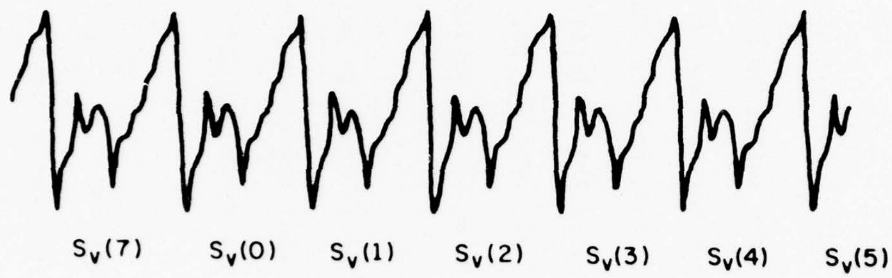


Figure 11. Serial Implementation

interpolate the waveform from the known samples. This is shown in Figure 12. Defining t_h as the time between sample points for harmonic h, one observes the following relations:

$$t_6 = (8/7)\,t_7 \qquad t_3 = (8/4)\,t_7$$
$$t_5 = (8/5)\,t_7 \qquad t_2 = (8/3)\,t_7$$
$$t_4 = (8/5)\,t_7 \qquad t_1 = (8/2)\,t_7$$
$$t_0 = (8/1)\,t_7$$

Linear interpolation is well suited to Burst processing. [6] If one slides a window between two Bursts, one observes an interpolation between the two known values. This results from the unary properties of Burst. Figure 12 demonstrates this interpolation. To perform the various convolutions in parallel, one need only use these interpolations as the necessary sample points which are passed through the H+1 transform units.

In this prototype, the serial approach was chosen for hardware implementation. It was felt that speech does exhibit enough redundancy to allow a serial computation. Hardware costs were also a factor in the design.

6.5 ASYNCHRONOUS PULSE MULTIPLIER

Harmonic self sampling requires a pitch synchronous, variable rate clock. The speech input must be sampled at a rate dependent on two parameters: T, the fundamental period of the speech; and h, the harmonic being computed. If the transform unit consists of 32 points, each 16 bits in length; 512 pulses must be inserted in the fundamental period. This is accomplished using the design shown in Figure 13.

Given pulses indicating the beginning of each fundamental period, the APM measures the present fundamental period and uses this value as

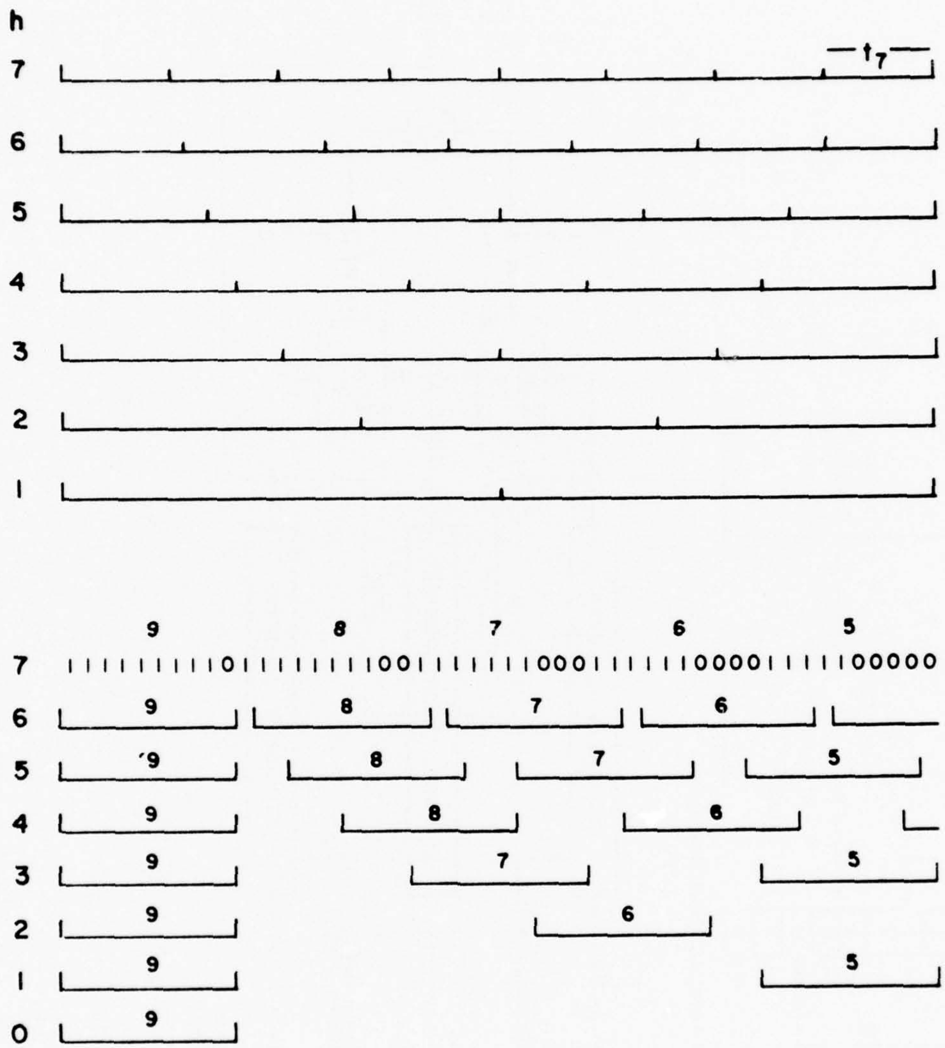


Figure 12. Burst Interpolation

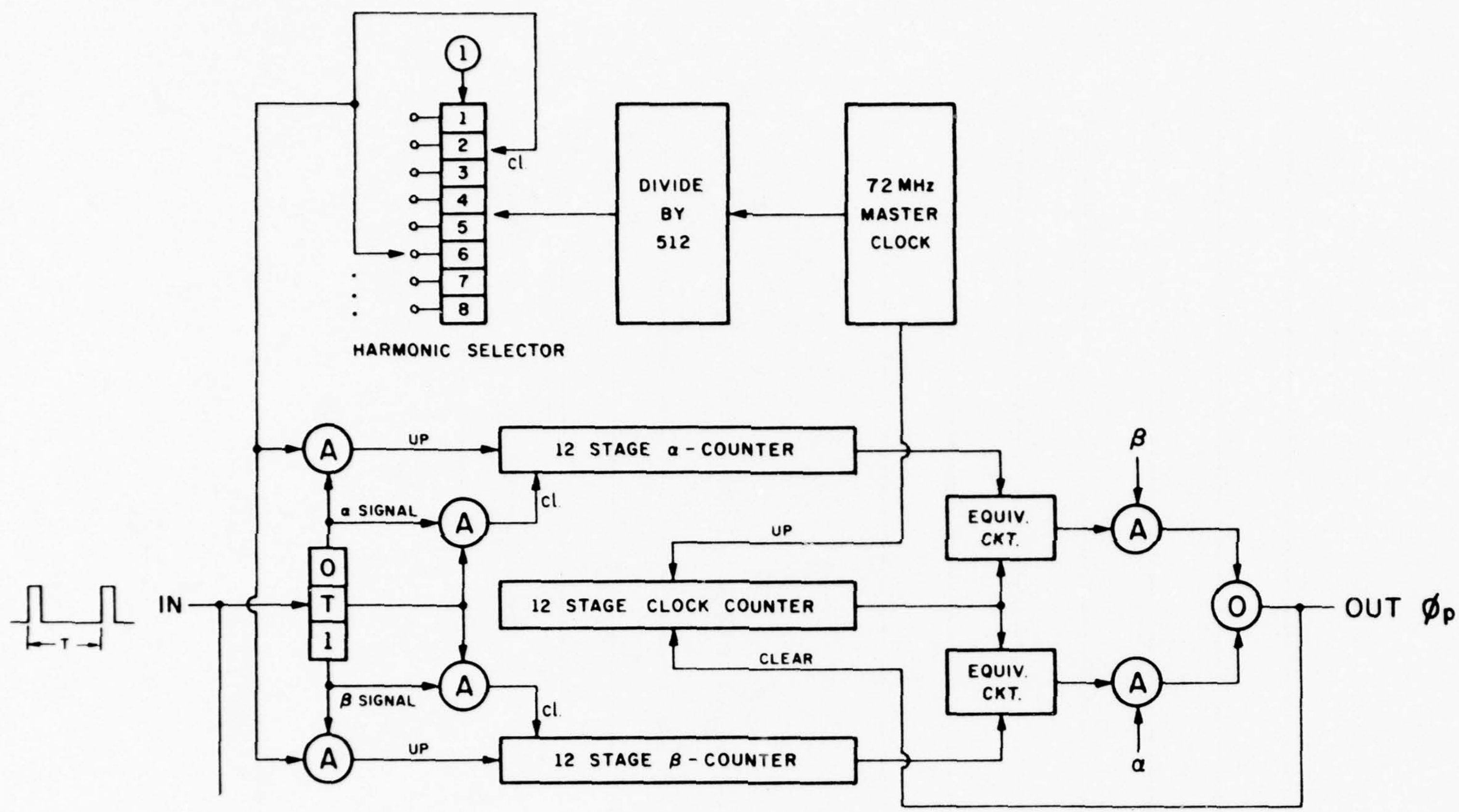


Figure 13. Asynchronous Pulse Multiplier

an estimate of the next period. Although not essential to the basic concept, this technique eliminates the necessity of delaying the input waveform by one period.

A standard method of frequency multiplication is to use a phase-locked loop on the harmonics of a clock. Due to the inertia present in such a method, it was rejected for a more direct approach. Using a high speed time reference, $\emptyset_s$, it is divided by 512 and by h, the harmonic to be computed next. This is used to drive the a-counter which measures the time T. During the next fundamental period, the a-counter is compared to a counter driven directly from $\emptyset_s$. Each time this clock counter equals the statisized a-counter, a pulse is generated and the clock counter is cleared. A pair of counters (a,b) are utilized so that one value is staticized while the next period is being computed.

The clock being implemented in 10000 series ECL circuitry, $\emptyset_s$ is 72 MHZ. Assuming a maximum fundamental period of 250 Hz, the fundamental period will be estimated to within 0.17% for h equal to 0, and to within 1.4% for h equal to 7.

A simulation study was undertaken to determine the accuracy of the period estimation for various fundamental periods. Choosing the period values at random, an estimate was computed for the eight harmonics. The relative error averaged over the eight results for each fundamental period tested is shown in Table 1. The error is obviously not a strictly increasing function of frequency. Since we are essentially performing an integer division of the period, there are values relative to $\emptyset_s$ which have varying truncation errors. This will account for the local discontinuities. It is noteworthy that for a fundamental as high at 1152 Hz, an average error of only 1.44% is observed.

Fundamental Period (Hz)	Average Error (%)
39.6	.06
79.9	.06
115.2	.13
144.0	.11
195.8	.24
246.2	.23
281.2	.30
303.8	.21
360.0	.37
426.2	.23
524.8	.59
600.5	.46
655.2	.75
720.0	.63
818.6	1.01
1023.1	1.4
1152.0	1.44

Table 1.

6.6 COEFFICIENT COMPUTATION

Given the results of the partial convolutions, the operations indicated in Eq. (3.7) must be performed. A block diagram describing the required operations is shown in Figure 14. The partial convolutions for a given fundamental period are summed together in a counter. The result is normalized with respect to h+1, the number of convolutions performed. The result must then be squared, summed with the corresponding sin/cos coefficient, and then the magnitude of the spectral line is produced by taking the square root.

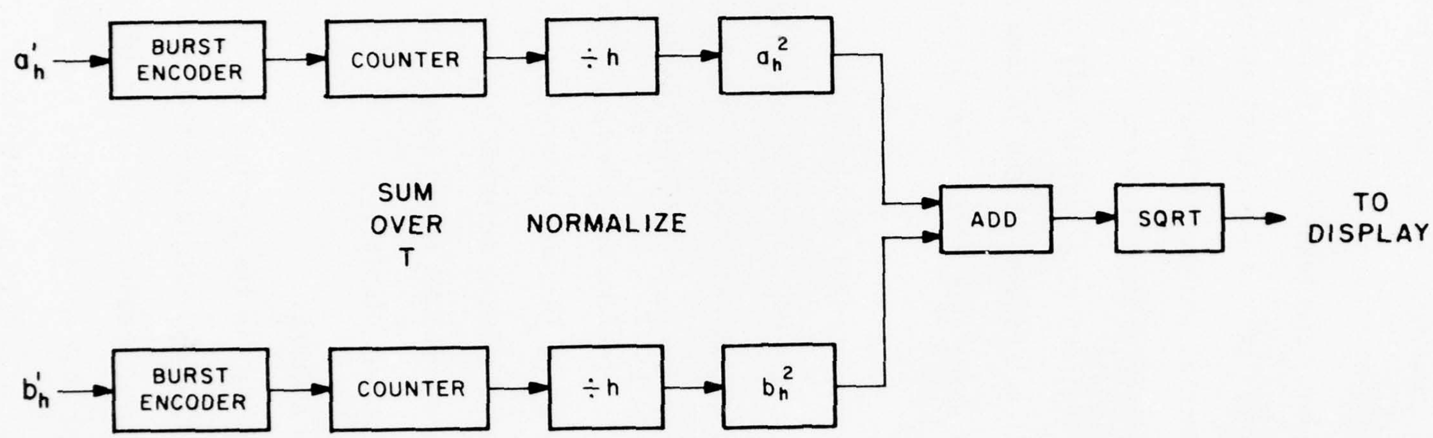


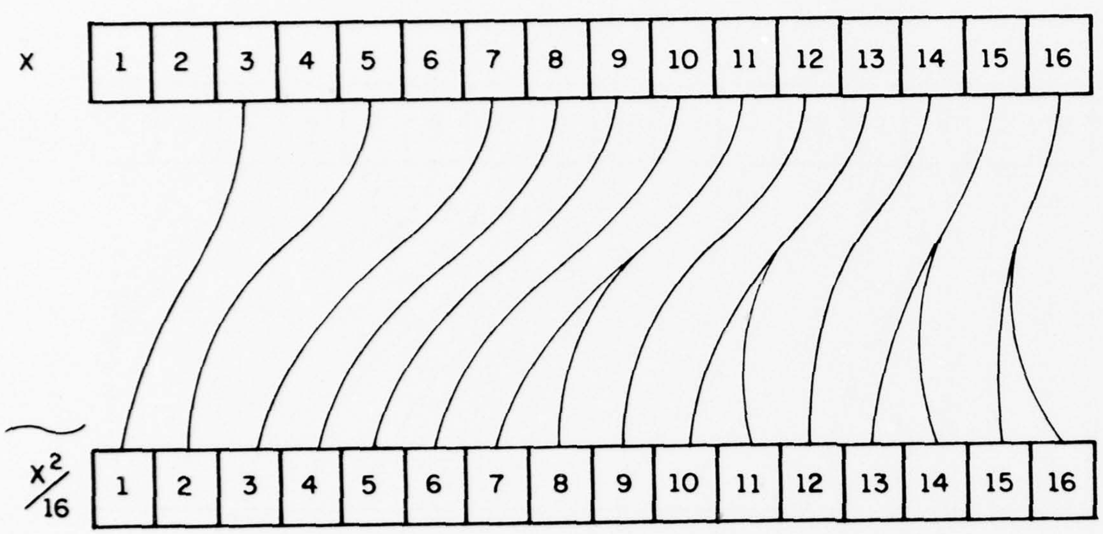
Figure 14. Coefficient Computation

The implementation of squaring, adding, and obtaining the square root is based on the unary properties inherent to Burst processing. Observing the value of a compacted Burst, the information is contained in the location of the 1-0 boundary. It is a positional attribute. This property lends itself to trivial function implementations. By correctly connecting the outputs of a Burst register to predetermined inputs of a second register, the contents of the receiving register will contain the Burst approximation to the function. Figures 15 and 16 show a squaring and square root implementation using this idea. One notices that appropriate scaling is necessary.

Burst addition may be implemented in several ways. [4,5] The method chosen consists of observing the odd pulses of the addend and the even pulses of the augend. The output, as shown in Figure 17, is a scaled approximation of the desired result. It is not possible to guarantee a compacted output, so one must perform compaction before further processing is allowed. Combining the three function generators to obtain the spectral coefficient, one arrives at the logic depicted in Figure 18. The squaring and addition connections are combined in one step. The final result is routed to the appropriate output display.

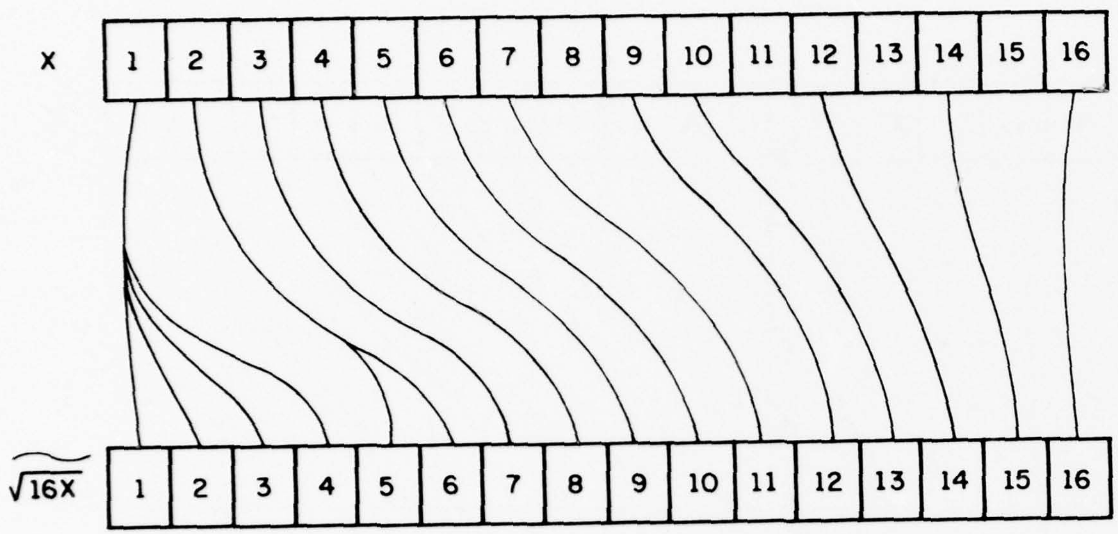
6.7 INCREASED COMPUTATION ACCURACY

The computations described in the previous section were implemented with 16 Burst digit accuracy. Assuming a uniform probability distribution for the possible input values, a mean square error of .065 was obtained for the squaring operation; .064 for the square root operation, and 0.25 for the addition. Using a uniform input distribution, the mean square error for the combined operations using 16 bits is .446. This should be regarded as an upper bound. It is generally accepted that speech exhibits a near Gaussian distribution. Such a distribution would effectively reduce this mean square error.



X	$\widetilde{X^2/16}$	E
0	0	0
1	0	.06
2	0	.25
3	1	.44
4	1	0.0
5	2	.44
6	2	.25
7	3	.44
8	4	0.0
9	5	.06
10	6	.25
11	8	.44
12	9	0.0
13	11	.44
14	12	.25
15	14	.06
16	16	0.0

Figure 15. Squaring Connections



X	$\widetilde{\sqrt{16X}}$	E
0	0	0.0
1	4	0.0
2	6	.34
3	7	.07
4	8	0.0
5	9	.06
6	10	.2
7	11	.42
8	11	.31
9	12	0.0
10	13	.35
11	13	.27
12	14	.14
13	14	.44
14	15	.03
15	15	.49
16	16	0.0

Figure 16. Square Root Connections

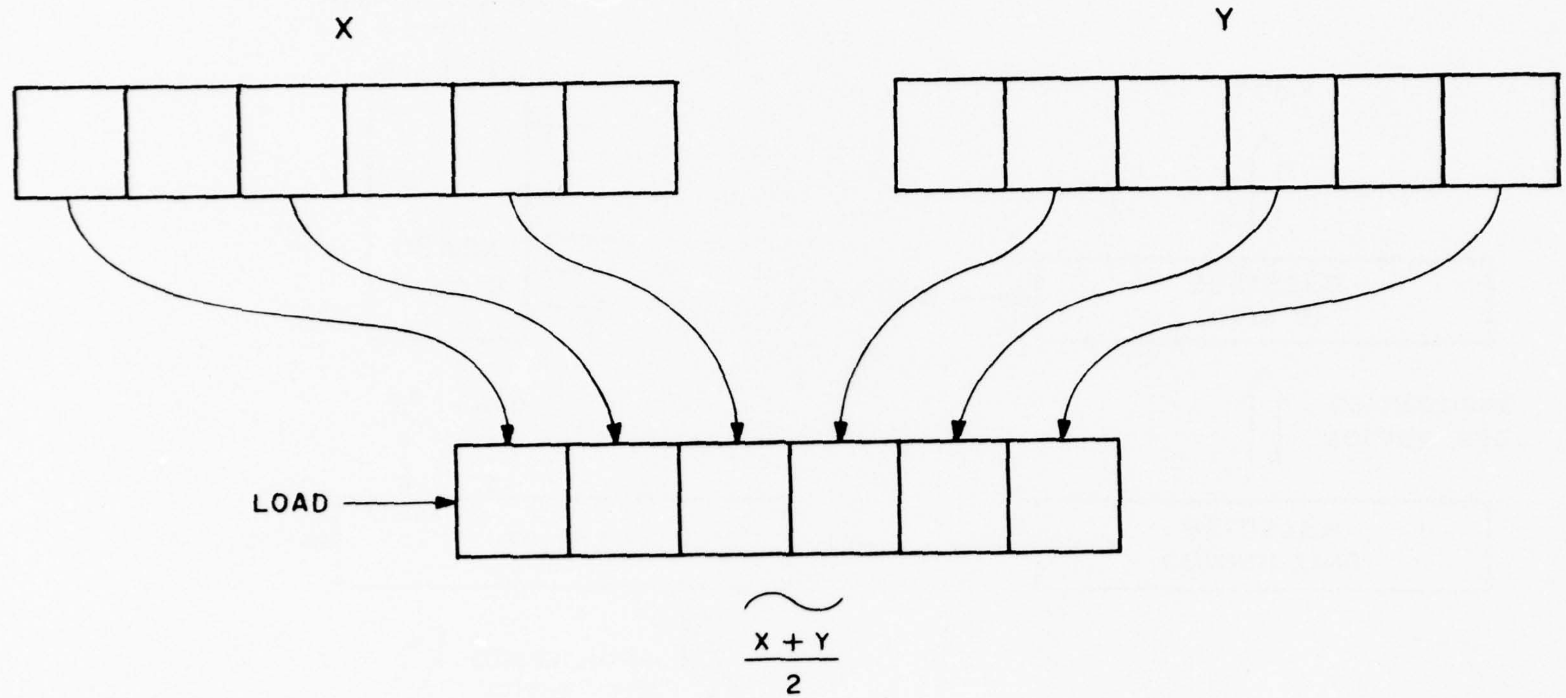


Figure 17. Burst Addition

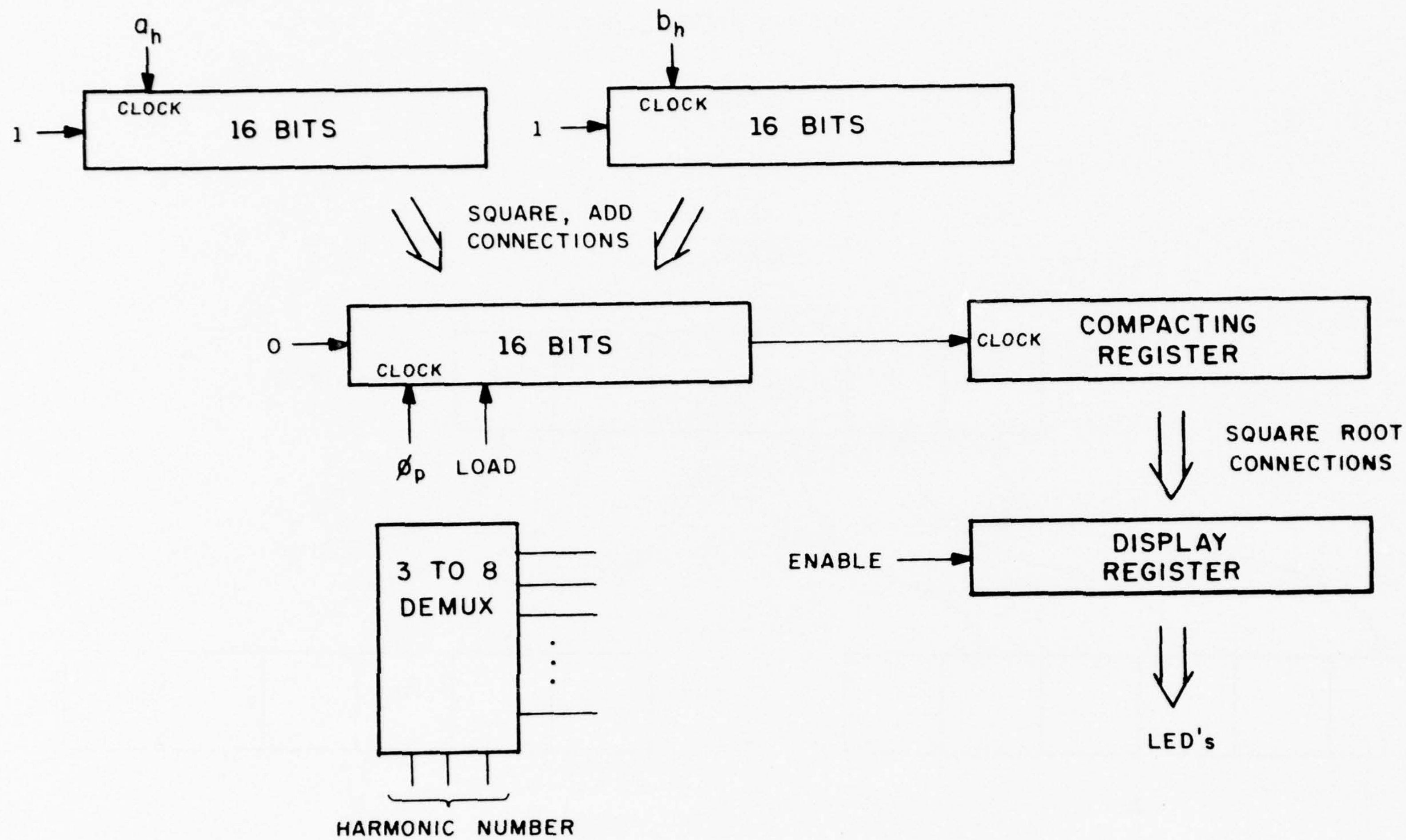


Figure 18. Coefficient Logic

Given any input distribution, one may reduce this computational error arbitrarily close to zero. Assuming a fixed number of bits for the input and output value, one may increase the number of bits used in the intermediate calculations. The principles described in the previous section remain valid. Figure 19 demonstrates this for the case of 10 bit input/output values and 20 bit function evaluations.

A simulation study has shown that the MSE decreases in an approximate exponential manner with increasing bit length. The results are shown graphically in Figure 20. Obviously we do not have a smooth function. One observes large discontinuities for lengths of 19, 24, and 43 bits. These values should be considered when making improvements.

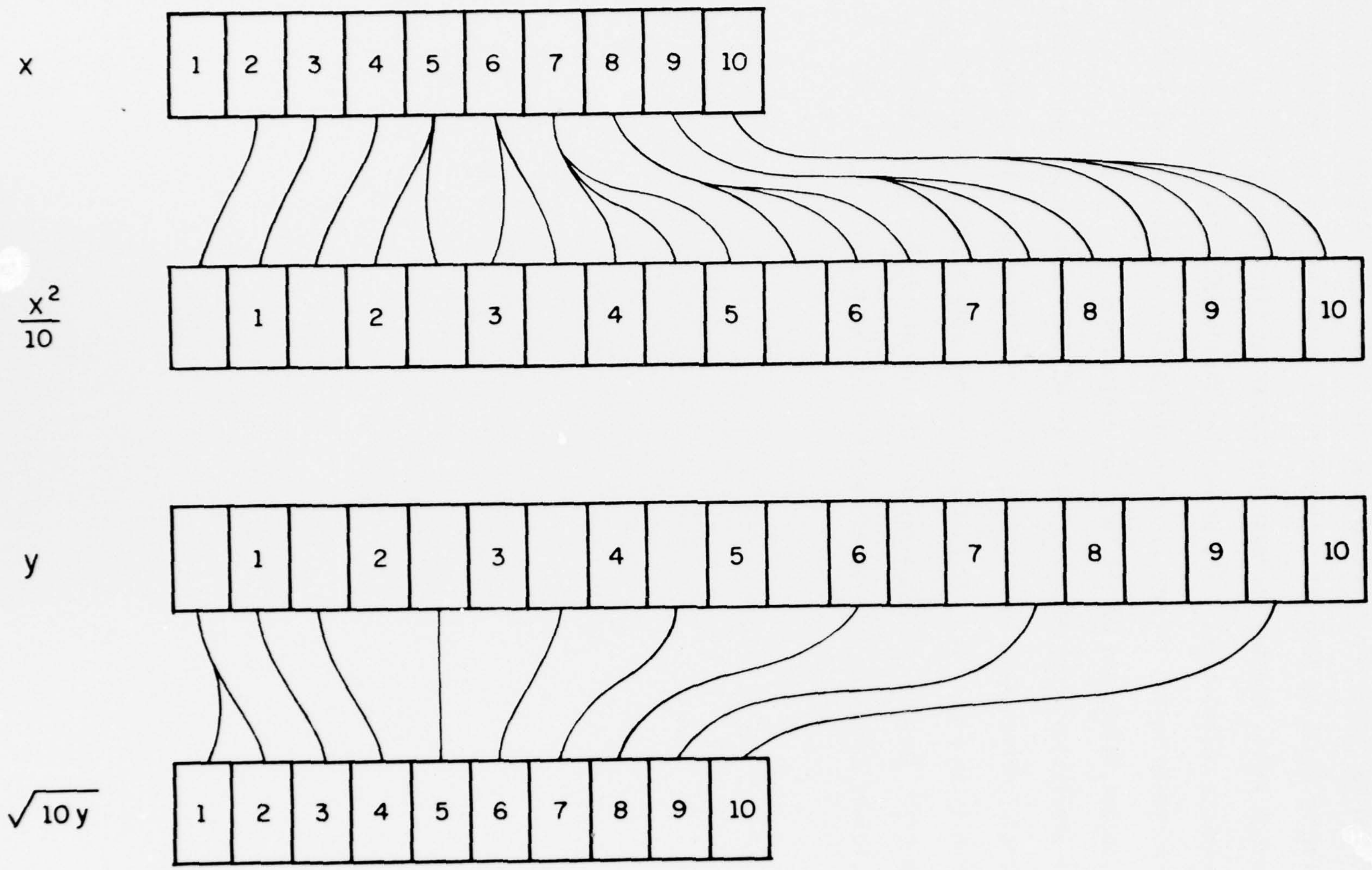


Figure 19. Increased Accuracy

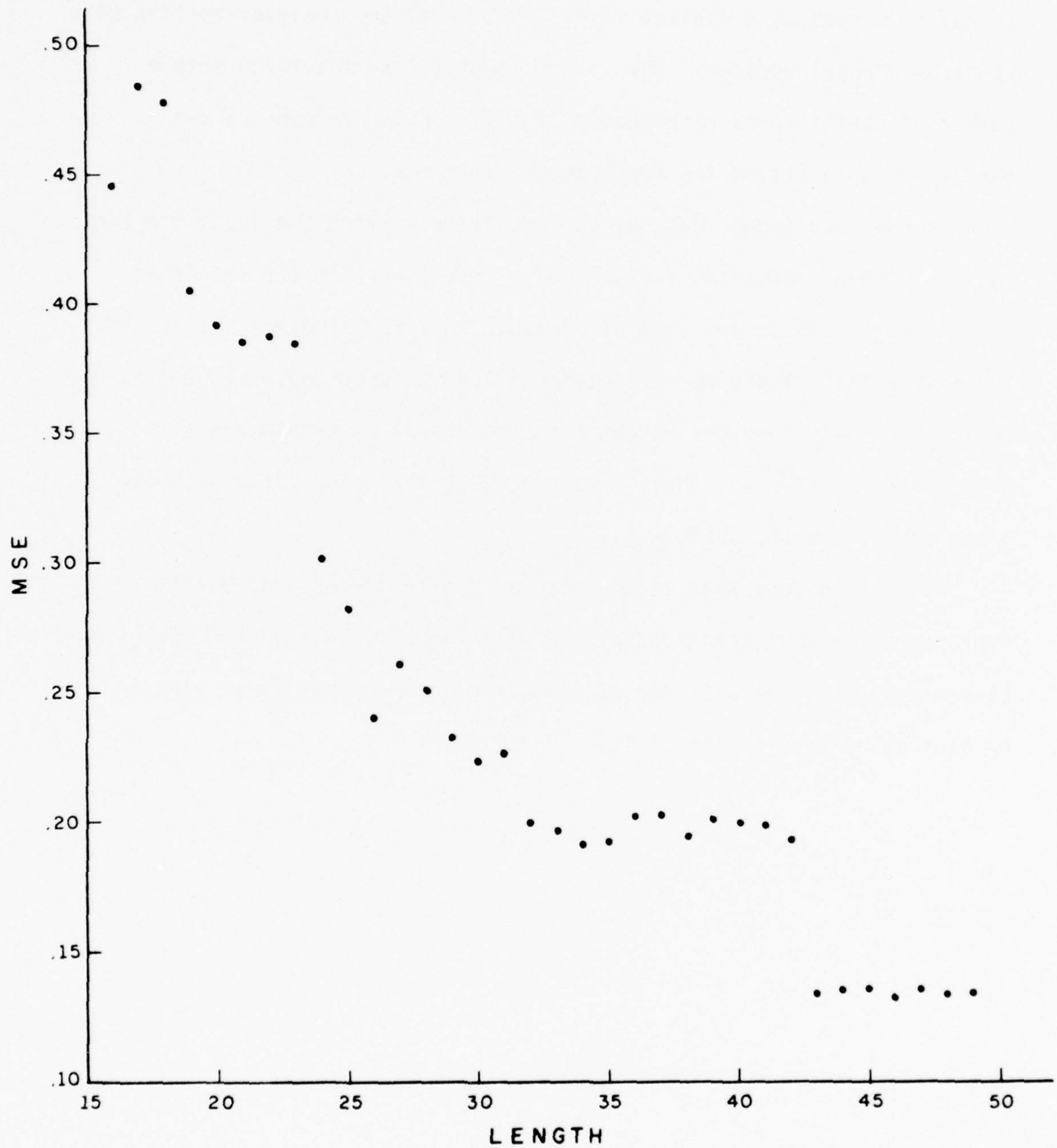


Figure 20. MSE vs Length

7. CONCLUSION

A real time speech analyzer using Burst Processing has been implemented. It may be viewed as a digital implementation of the analyzer portion of a spectrum channel vocoder. The speech input is parameterized into a number of coefficients representing the spectral envelope and one parameter representing the fundamental frequency.

It has been shown that, by appropriately varying the input sampling rate, a single tranversal filter may be used to obtain the required coefficients. Thus, the idea of Harmonic Self-sampling was introduced. The flexibility of the Burst implementation is demonstrated by the fact that essentially the same hardware can be used to generate other orthogonal transforms. Thus, Hadamard, Chebyshev, and Karhunen-Loeve transforms are also possible.

We may conclude that in the area of speech processing, Burst representation does indeed provide a promising alternative to conventional binary systems. The analyzer described is an important first step in demonstrating this applicability.

REFERENCES

[1] Campanella, S. and Robinson, G., "A Comparison of Orthogonal Transformations for Digital Signal Processing," IEEE Transactions on Communication Technology, COM-19, December 1971.

[2] Robinson, G. and Granger, R., "Fast Fourier Transform Speech Compression," Proceedings of the 1970 IEEE International Conference on Communications, paper 26-5, June 1970.

[3] Poppelbaum, W. J., Appendix to "A Practicability Program in Stochastic Processing," Department of Computer Science, University of Illinois, March 1974.

[4] Bracha, E., "Burstcalc (A Burst Calculator)," Report UIUCDCS-R-75-769, Department of Computer Science, University of Illinois, October 1975.

[5] Taylor, G., "An Analysis of Burst Encoding Methods," Report UIUCDCS-R-75-770, Department of Computer Science, University of Illinois, December 1975.

[6] Mohan, P., "The Application of Burst Processing to Digital FM Receivers," Report UIUCDCS-R-76-780, Department of Computer Science, University of Illinois, January 1976.

[7] Pleva, R., "A Microprocessor-Controlled Interface for Burst Processing," Report UIUCDCS-R-76-812, Department of Computer Science, University of Illinois, July 1976.

[8] Wolff, M., "Transmission of Analog Signals Using Burst Techniques," Report UIUCDCS-R-77-838, Department of Computer Science, University of Illinois, January 1977.

[9] Cohen, J., "An Automatic Process for Pitch Period Extraction of Speech," Masters Thesis, Department of Electrical Engineering, University of Illinois, 1970.

[10] Noll, A., "Short Time Spectrum and Cepstrum Techniques for Vocal Pitch Detection," Journal Acoustical Society of America, Vol. 36, 1964.

[11] Flanagan, J., "Speech Analysis Synthesis and Perception," Academic Press, Inc., New York, 1965.

[12] Bergland, G., "Fast Fourier Transform Hardware Implementations--An Overview," IEEE Transactions on Audio and Electroacoustics, Vol. AU-17, June 1969.

[13] Liu, J., "Application of Burst Code and Burst Processing Techniques in Communication Systems," Internal Memo, May 1977.

[14] Tolstov, G., "Fourier Series," Prentice-Hall, Inc., New Jersey, 1962.

[15] Schroeder, M., "Vocoders: Analysis and Synthesis of Speech," Proceedings of the IEEE, Vol. 54, No. 5, May 1966.

[16] Poppelbaum, W. J., "Statistical Processors," Advances in Computers, Vol. 14, 1976.

[17] Poppelbaum, W. J., "Application of Stochastic and Burst Processing to Communication and Computing Systems," Department of Computer Science, University of Illinois, March 1976.

APPENDIX

CIRCUIT DRAWINGS

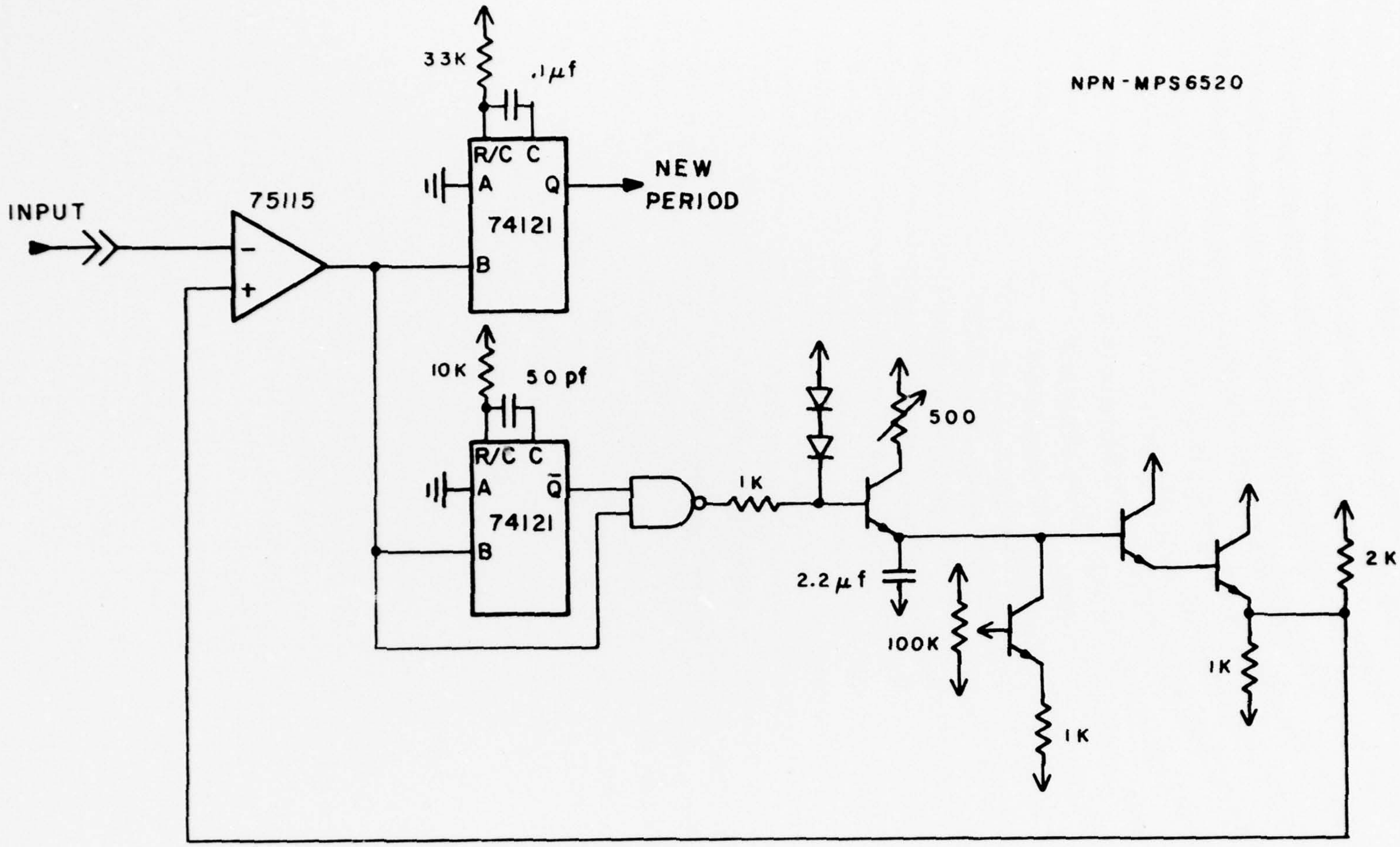


Figure A1. Pitch Period Detection

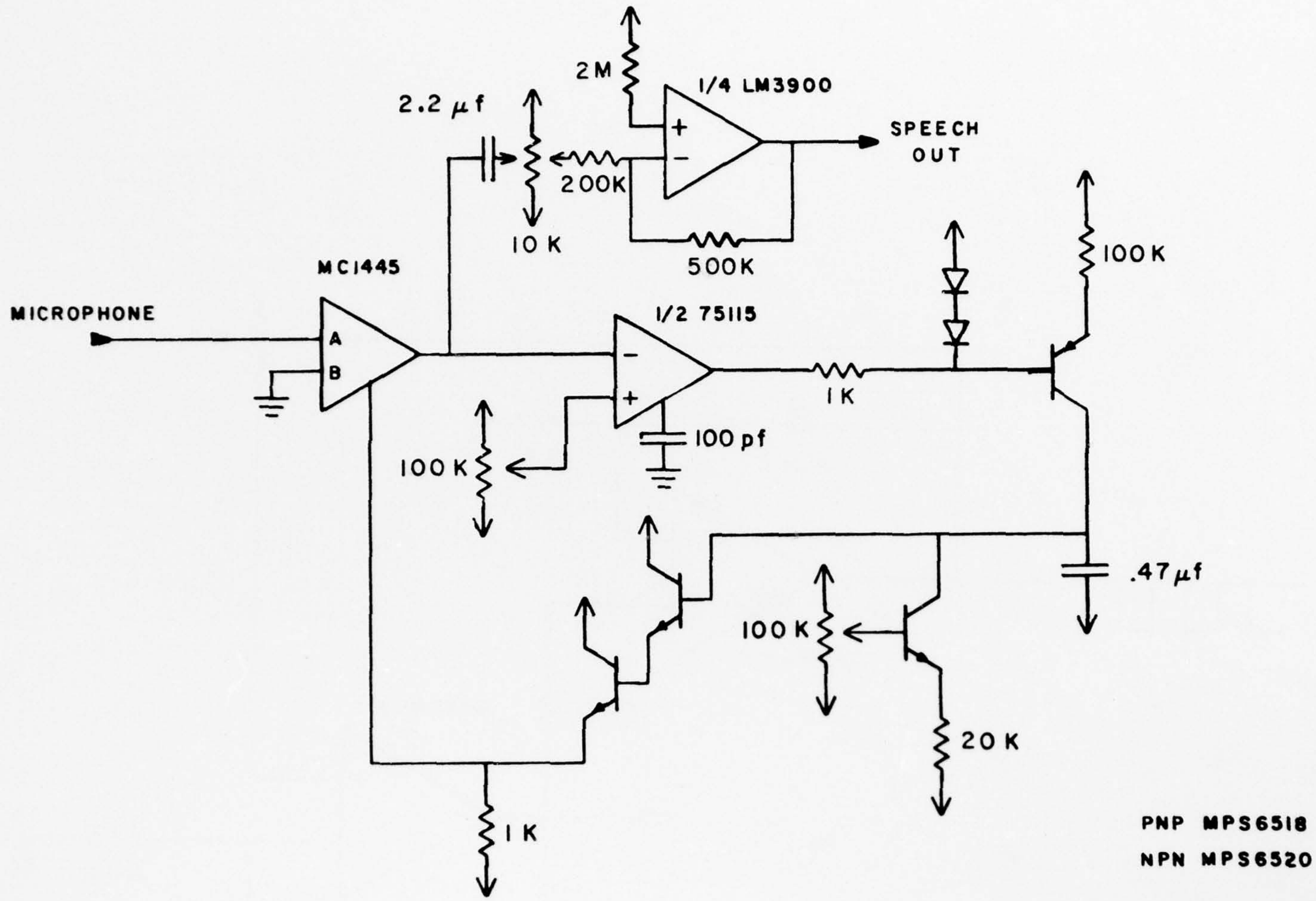


Figure A2. Automatic Gain Control

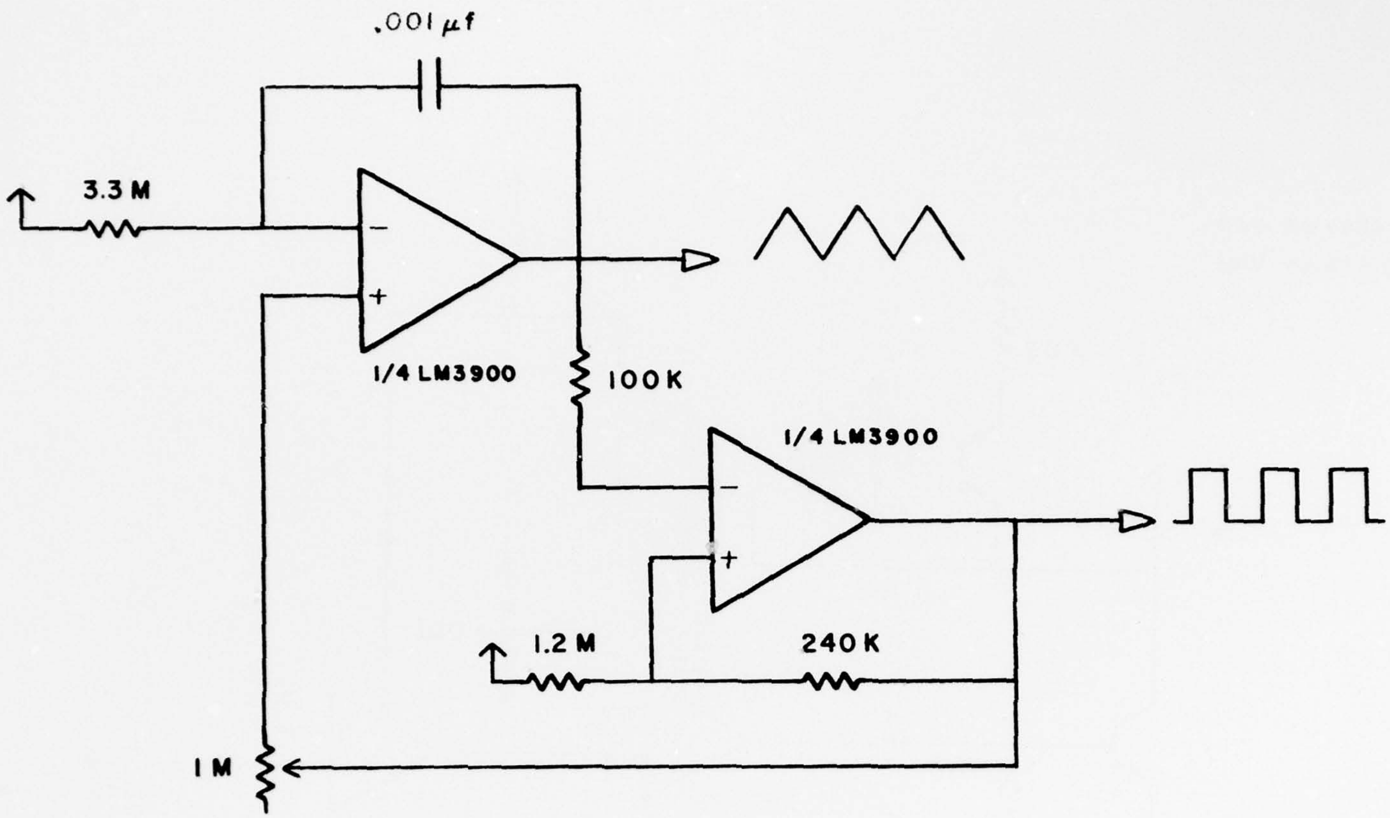


Figure A3. Waveform Generator

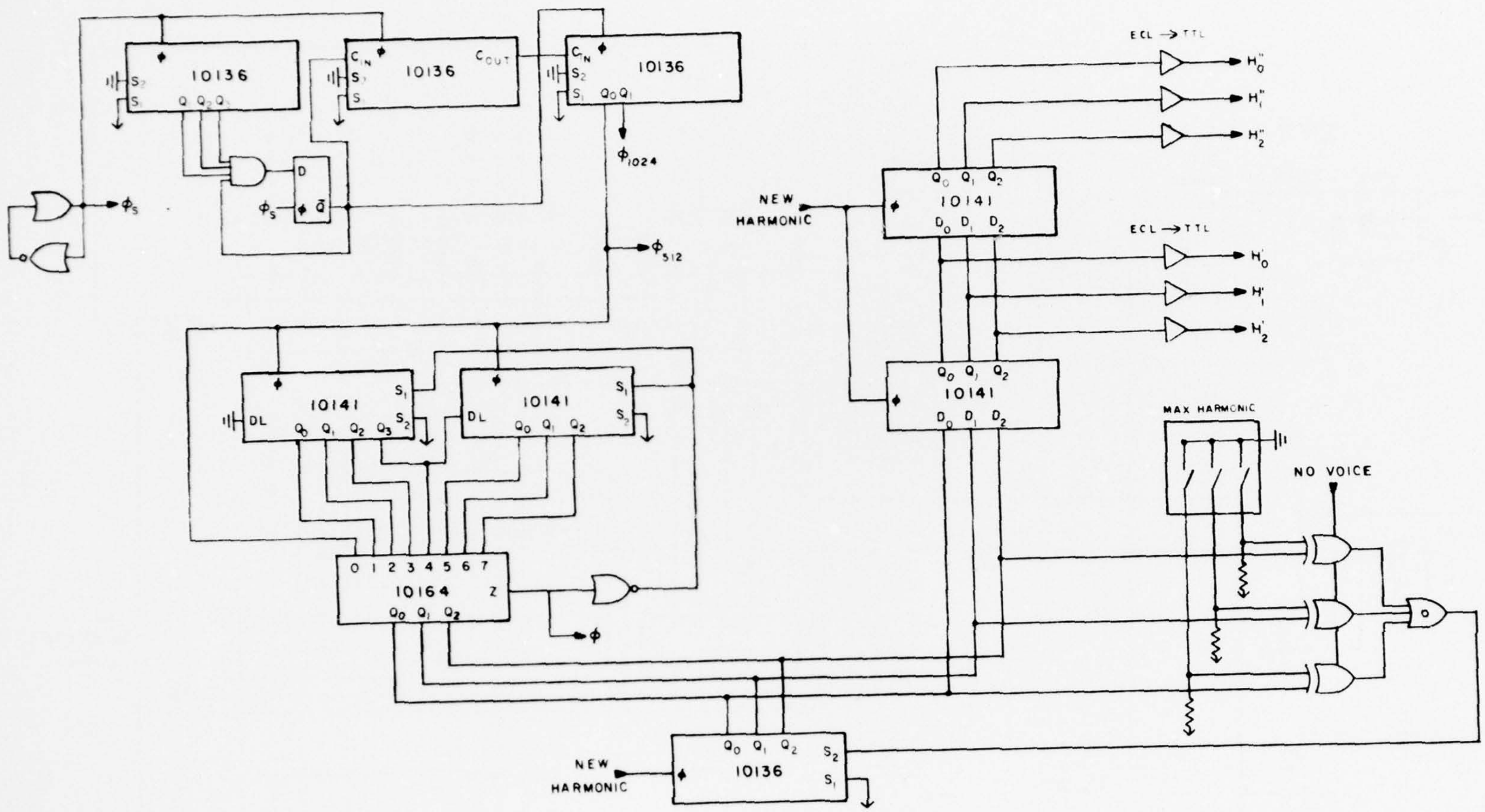


Figure A4. Clock Count Down

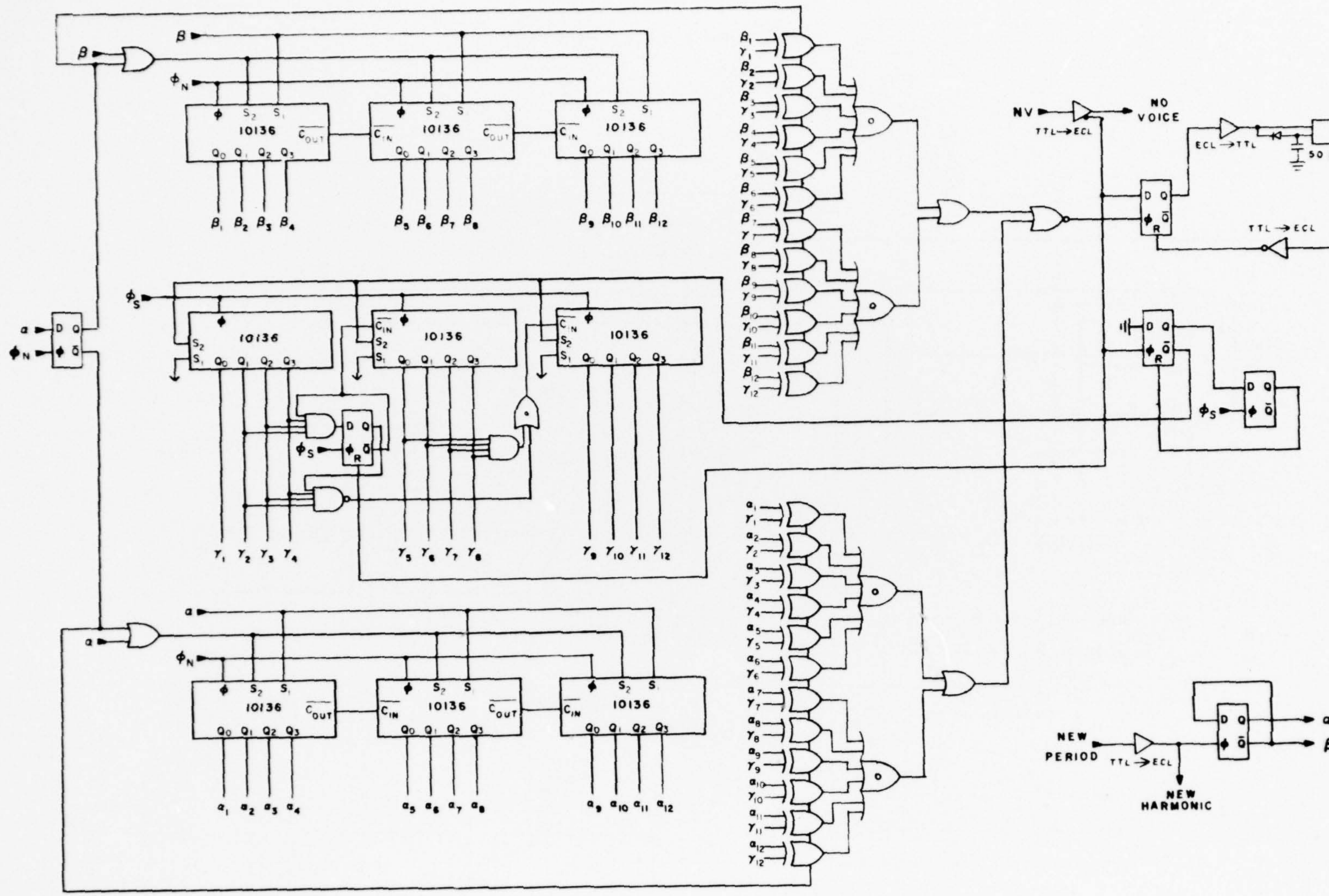


Figure A5. Pulse Multiplier

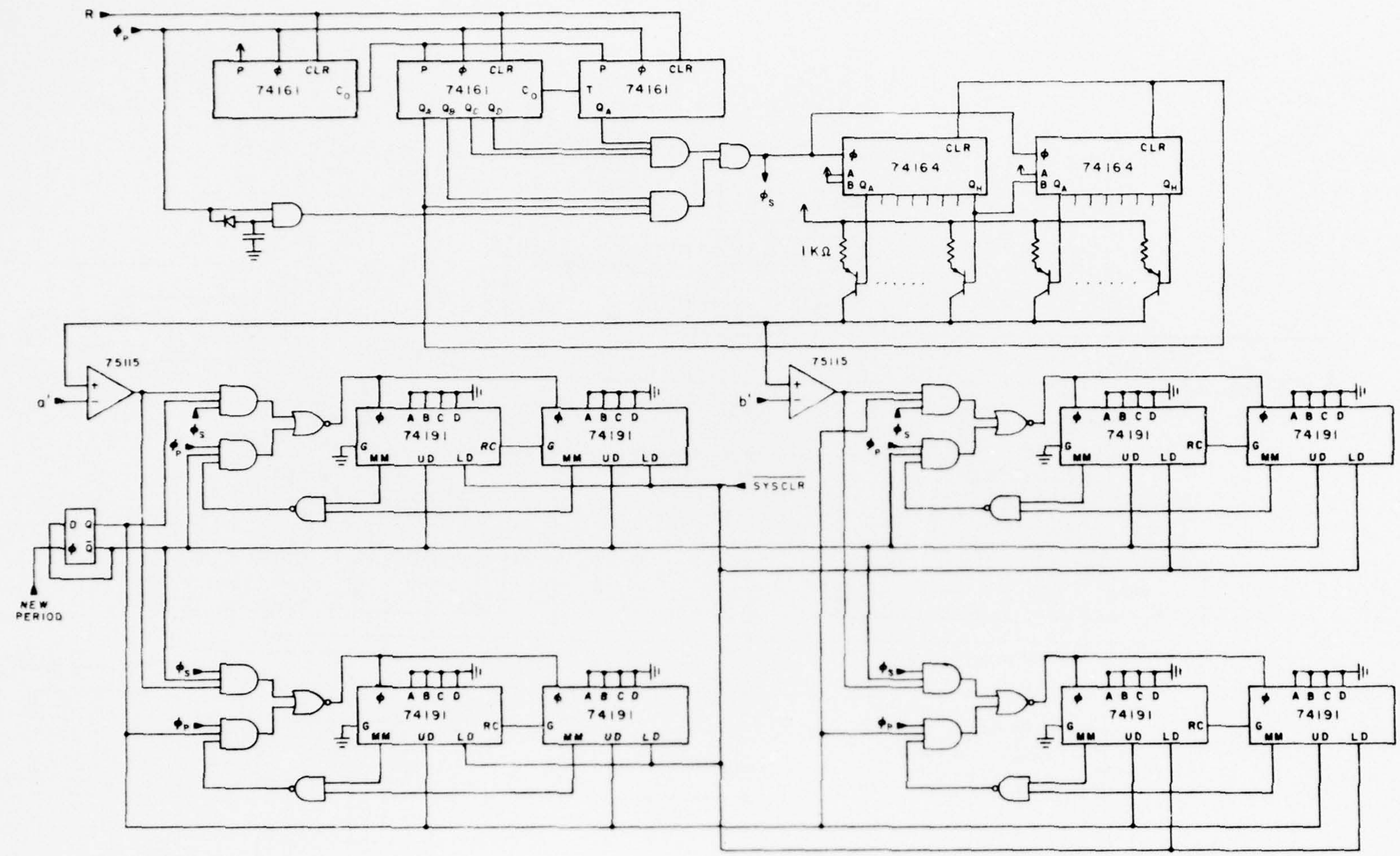


Figure A6. Coefficient Accumulators

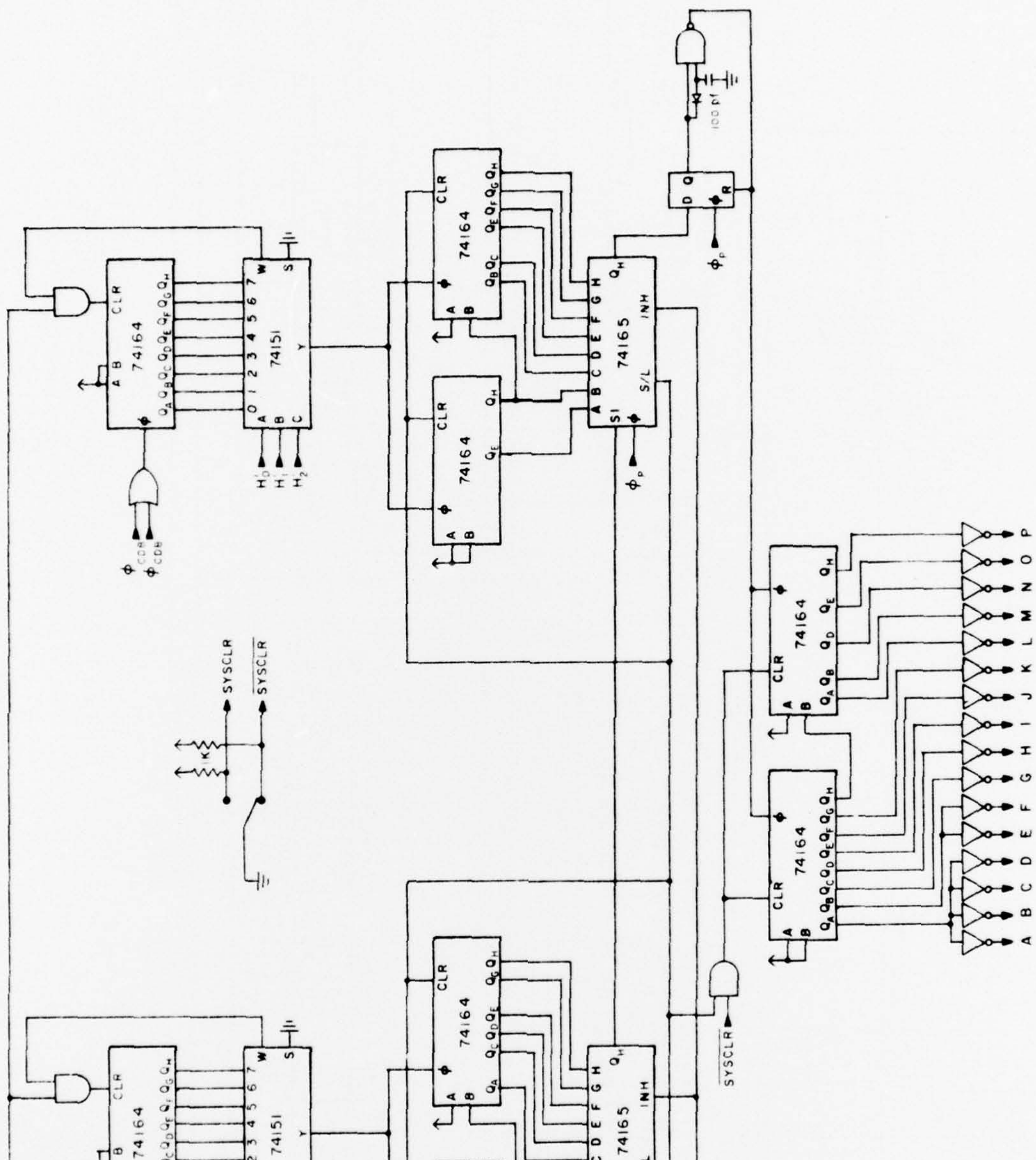


Figure A7. Coefficient Computation

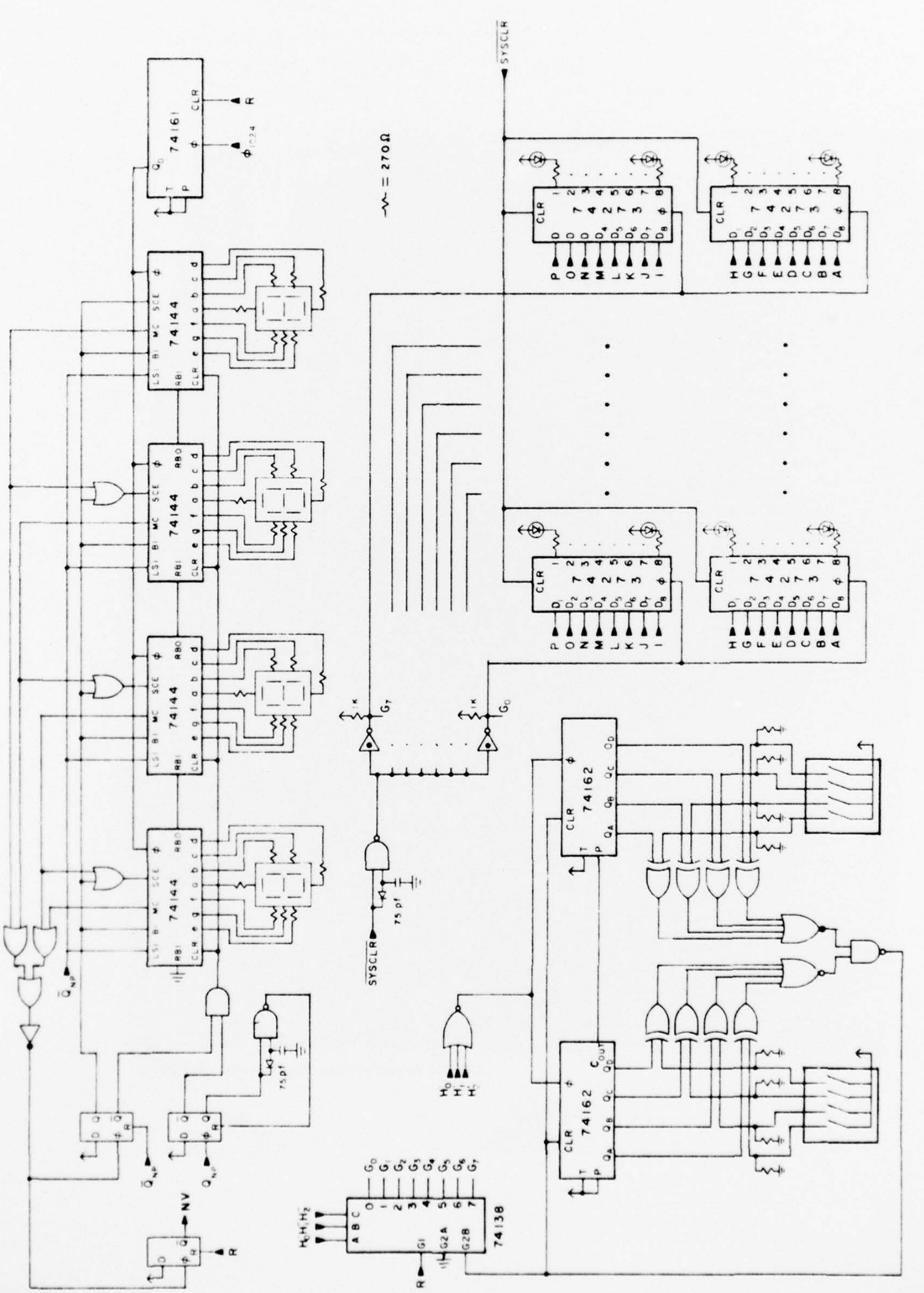


Figure A8. Output Display

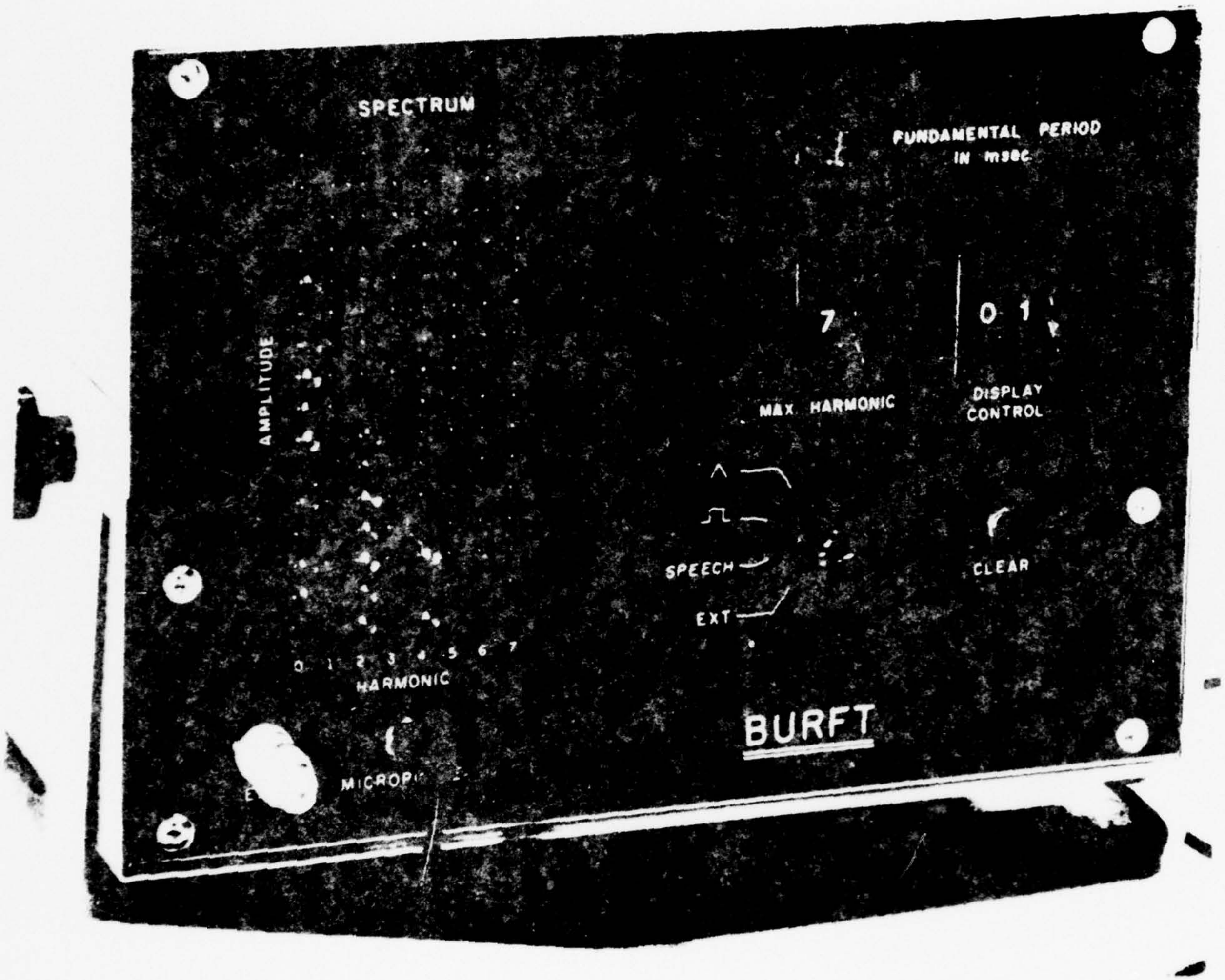


Figure A9. Analyzer Prototype

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER UIUCDCS-R-77-870	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) APPLICATION OF BURST PROCESSING TO THE SPECTRAL DECOMPOSITION OF SPEECH		5. TYPE OF REPORT & PERIOD COVERED M.S. Thesis/June 1977
		6. PERFORMING ORG. REPORT NUMBER UIUCDCS-R-77-870
7. AUTHOR(s) Christ John Xydes		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0982
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Computer Science University of Illinois at Urbana-Champaign Urbana, Illinois 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Code 437 Arlington, Virginia 22217		12. REPORT DATE
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Release Unlimited
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Burst Processing
Harmonic Self Sampling
Vocoder
Fourier Transform
Block Sum Register
Transversal Filter

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The application of Burst Processing to the problem of spectral decomposition of speech is discussed. It is shown that such a representation provides a viable alternative to conventional speech analyzers. A specific Burst implementation is presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BIBLIOGRAPHIC DATA SHEET	1. Report No. UIUCDCS-R-77-870	2.	3. Recipient's Accession No.
4. Title and Subtitle APPLICATION OF BURST PROCESSING TO THE SPECTRAL DECOMPOSITION OF SPEECH			5. Report Date June 1977
			6.
7. Author(s) CHRIST JOHN XYDES			8. Performing Organization Rept. No. UIUCDCS-R-77-870
9. Performing Organization Name and Address Department of Computer Science University of Illinois at Urbana-Champaign Urbana, IL 61801			10. Project/Task/Work Unit No.
			11. Contract/Grant No. N00014-75-C-0982
12. Sponsoring Organization Name and Address Office of Naval Research Code 437 Arlington, VA 22217			13. Type of Report & Period Covered M.S. Thesis
			14.
15. Supplementary Notes			

16. Abstracts

The application of Burst Processing to the problem of spectral decomposition of speech is discussed. It is shown that such a representation provides a viable alternative to conventional speech analyzers. A specific Burst implementation is presented.

17. Key Words and Document Analysis. 17a. Descriptors

Burst Processing
Harmonic Self Sampling
Vocoders
Fourier Register
Transversel Filter

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages
Release Unlimited	20. Security Class (This Page) UNCLASSIFIED	22. Price